Álgebra Linear
TEXTO EM FASE DE PREPARAÇÃO

Jones Colombo José Koiller

Departamento de Análise
Instituto de Matemática e Estatística
Universidade Federal Fluminense

Conteúdo

Capítulo 1

Sistemas Lineares e Escalonamento de Matrizes

1.0 Introdução

O leitor, ou a leitora, desde o ensino fundamental e médio, já deve ter se deparado muitas vezes com sistemas de equações do tipo

$$\underset{\text{Sistema (a)}}{\begin{cases} x_1 - 3x_2 = -1 \\ 2x_1 + x_2 = 5, \end{cases}} \qquad \underset{\text{Sistema (b)}}{\begin{cases} 6x_1 + 3x_2 = 9 \\ 2x_1 + x_2 = 5, \end{cases}} \quad \text{ou} \quad \underset{\text{Sistema (c)}}{\begin{cases} 6x_1 + 3x_2 = 15 \\ 2x_1 + x_2 = 5. \end{cases}}$$

Esses são exemplos de *sistemas lineares*, objeto de estudo deste capítulo, que serão úteis durante todo o curso.

Você deve ter aprendido, no ensino fundamental, ao menos dois métodos para resolver os sistemas acima. Um deles é o *método da substituição.* No sistema (a), por exemplo, podemos "isolar" a variável x_1 na primeira equação, obtendo, assim, $x_1 = 3x_2 - 1$. Em seguida, substituimos x_1 por $3x_2 - 1$ na segunda equação, obtendo $2(3x_2 - 1) + x_2 = 5$. Resolvendo esta equação, encontramos $x_2 = 1$. Agora, basta colocar $x_2 = 1$ em $x_1 = 3x_2 - 1$ para concluirmos que $x_1 = 2$. Portanto, a única solução do sistema (a) é dada por $x_1 = 2$ e $x_2 = 1$.

O segundo método que você deve conhecer é o *método da eliminação.* Multiplicando a primeira equação do sistema (a) por 2, obtemos $2x_1 - 6x_2 = -2$. Subtraindo esta nova equação da segunda equação do sistema, obtemos

$$\begin{array}{rrcr} & 2x_1 + x_2 & = & 5 \\ -\big(& 2x_1 - 6x_2 & = & -2\ \big) \\ \hline & 7x_2 & = & 7, \end{array}$$

ou seja, $x_2 = 1$, como havíamos obtido. Agora, podemos substituir $x_2 = 1$ em qualquer uma das equações do sistema (a) para achar $x_1 = 2$.

Os métodos que acabamos de recapitular são satisfatórios para tratar sistemas pequenos, como os acima, que têm apenas duas equações e duas variáveis. Mas imagine a dor-de-cabeça que seria aplicá-los a sistemas com cinco, seis, ou até mais variáveis e equações! Precisamos desenvolver um método que nos permita lidar, mais facilmente, com sistemas deste porte, pois teremos que fazê-lo, muitas vezes, ao longo do curso.

Um dos objetivos deste capítulo é estudar um procedimento, chamado de *eliminação Gaussiana* ou *escalonamento*, para resolver sistemas lineares. Esse procedimento é, essencialmente, o método da eliminação acrescido de uma sistemática para a organização do trabalho. A sistemática é importante para reduzir o esforço e a chance de cometermos erros de cálculo.

Observação
O desenvolvimento de um método sistemático para resolver sistemas lineares é importante por diversos motivos. Talvez, o motivo principal seja o seguinte: este é o primeiro passo para se escrever um programa de computador que realize esta tarefa. Muitos problemas oriundos da modelagem matemática de sistemas físicos, biológicos, econômicos e de outras áreas — incluindo importantes aplicações a todas as áreas da engenharia — podem envolver milhares, centenas de milhares, milhões de variáveis e equações. Para resolver tais sistemas, é necessário usar recursos computacionais, evidentemente.

Outro objetivo deste capítulo é obter uma caracterização de sistemas lineares que têm uma única solução, sistemas que têm mais de uma solução, e sistemas que não têm solução alguma.

O sistema (a), por exemplo, tem uma única solução ($x_1 = 2$, $x_2 = 1$), como vimos acima. Esse fato possui uma interpretação geométrica simples. Lembre-se de que uma equação do tipo $ax_1 + bx_2 = c$ representa uma *reta* no plano (x_1, x_2). As retas correspondentes às equações do sistema (a) se intersectam em um único ponto (a saber, o ponto $(x_1, x_2) = (2, 1)$), como ilustra a figura 1.1(a). Este é o único ponto do plano (x_1, x_2) que pertence a ambas as retas, ou seja, é o único ponto que satisfaz simultaneamente as duas equações do sistema (a).

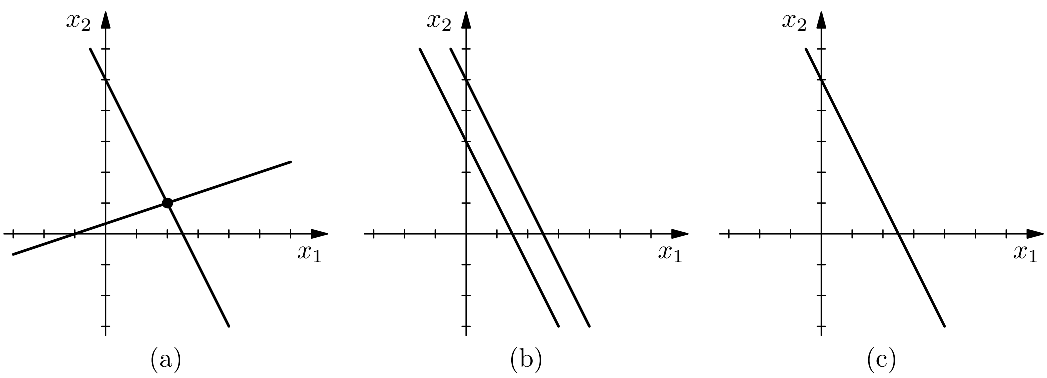


Figura 1.1: Interpretação geométrica dos sistemas (a), (b) e (c).

Você deve verificar, no entanto, que o sistema (b) não tem solução alguma, enquanto que o sistema (c) tem uma infinidade de soluções. Geometricamente, as equações do sistema (b) representam retas paralelas, que não se intersectam em ponto algum (figura 1.1(b)). Por outro lado, verifique que ambas as equações do sistema (c) representam *a mesma reta* (ilustrada na figura 1.1(c)). Dessa forma, a interseção dessa reta com ela própria é, novamente, ela própria!

Essa discussão será retomada nas seções 1.7 e 1.8. Veremos que o “cenário” exibido pelos sistemas (a), (b) e (c) permanece válido em um contexto mais abrangente: um sistema linear qualquer, e de qualquer tamanho, ou tem exatamente uma solução, ou nenhuma, ou uma infinidade delas.

1.1 Sistemas de equações lineares

Uma **equação linear** nas variáveis $x_1, x_2, \dots, x_n$ é uma equação que pode ser escrita na forma

$$c_1x_1 + c_2x_2 + \cdots + c_nx_n = b, \tag{1.1}$$

onde $c_1, c_2, \dots, c_n$ e b são números reais ou complexos.[1] Os números $c_1, \dots, c_n$ são chamados de **coeficientes** das variáveis $x_1, \dots, x_n$, respectivamente. Já b é chamado de **termo independente** da equação (1.1).

A equação

$$2x_1 - 3x_2 + 4 = x_1 - 5x_3,$$

por exemplo, é linear, pois pode ser reescrita como

$$x_1 - 3x_2 + 5x_3 = -4,$$

que está na forma da equação (1.1). Os coeficientes de x_1, x_2 e x_3 são 1, -3 e 5, respectivamente. O termo independente é -4. Já as equações

$$x_1^2 + 3x_2 = 5, \qquad x_1 + \sqrt{x_2} = 7 \qquad \text{e} \qquad 3x_1 + x_1x_2 - x_2 = 2$$

não são lineares. A primeira equação tem o termo não-linear x_1^2, a segunda tem o termo $\sqrt{x_2}$, e a terceira tem o termo x_1x_2.

Um **sistema de equações lineares**, ou, simplesmente, um **sistema linear**, é um conjunto (finito) de equações lineares envolvendo as mesmas variáveis. Cada um dos sistemas (a), (b) e (c) da introdução, por exemplo, é um sistema de duas equações lineares envolvendo as duas variáveis x_1 e x_2. Já o sistema linear

$$\begin{cases} x_1 - 2x_2 + x_3 = 0 \\ 2x_1 - 2x_2 - 6x_3 = 8 \end{cases} \tag{1.2}$$

tem duas equações envolvendo três variáveis: x_1, x_2 e x_3.

Uma **solução** de um sistema linear nas variáveis $x_1, x_2, \dots, x_n$ é uma lista de números $(s_1, s_2, \dots, s_n)$ que torna verdadeira cada igualdade do sistema quando substituímos $x_1, x_2, \dots, x_n$ pelos números $s_1, s_2, \dots, s_n$. Por exemplo, $(8, 4, 0)$ é uma solução de (1.2), pois $8-2\cdot4+0 = 0$ e $2\cdot8-2\cdot4-6\cdot0 = 8$. Logo, a primeira equação do sistema reduz-se a $0 = 0$, e a segunda, a $8 = 8$. Verifique que $(15, 8, 1)$ também é uma solução. Há uma infinidade de soluções além destas duas. Procure achar mais algumas! Já a lista $(2, 1, 0)$ *não* é uma solução do sistema (1.2), pois, apesar de reduzir a primeira equação a $0 = 0$ (o que é verdadeiro), ela reduz a segunda equação a $2 = 8$, o que é falso.

[1] Neste curso, consideraremos quase exclusivamente o caso real.

Definimos o **conjunto-solução** de um sistema linear como o conjunto de todas as suas soluções. Podemos reformular as afirmativas do parágrafo anterior dizendo que as listas $(8, 4, 0)$ e $(15, 8, 1)$ *pertencem ao conjunto-solução* do sistema (1.2), e que a lista $(2, 1, 0)$ *não pertence ao conjunto-solução.*

Se um sistema tem ao menos uma solução, dizemos que ele é **possível**. Do contrário, dizemos que é **impossível**. Os sistemas (a) e (c) da introdução são possíveis, enquanto o sistema (b) é impossível. O conjunto-solução de um sistema impossível é o conjunto vazio.

Encerramos esta pequena seção com uma última definição. Dizemos que dois sistemas são **equivalentes** quando eles têm o mesmo conjunto-solução. Os "sistemas" $2x_1 = 3$ e $4x_1 = 6$, por exemplo, são evidentemente equivalentes.

1.2 Notação matricial e o método de escalonamento: primeiro exemplo

Vamos introduzir uma notação, usando *matrizes*, que nos permita escrever sistemas lineares de forma mais econômica. Lembre-se, do ensino médio, que uma matriz é simplesmente um arranjo retangular de números. Estes são chamados de *entradas* ou *elementos* da matriz. Considere, por exemplo, o sistema

$$\begin{cases} x_1 - 4x_2 + x_3 = 0 \\ 3x_2 - 6x_3 = -3 \\ -3x_1 + 14x_2 - 6x_3 = 2. \end{cases} \tag{1.3}$$

Dizemos que

$$\begin{bmatrix} 1 & -4 & 1 \\ 0 & 3 & -6 \\ -3 & 14 & -6 \end{bmatrix}$$

é a **matriz de coeficientes** associada ao sistema (1.3), e que

$$\begin{bmatrix} 1 & -4 & 1 & 0 \\ 0 & 3 & -6 & -3 \\ -3 & 14 & -6 & 2 \end{bmatrix} \tag{1.3$'$}$$

é a **matriz completa** associada ao sistema. Outra maneira seria dizer que a matriz completa (1.3′) *representa* o sistema (1.3). Observe que a entrada na primeira coluna e segunda linha da matriz é zero, pois x_1 não aparece na segunda equação de (1.3), ou seja, seu coeficiente é zero.

Para escrevermos a matriz completa de um sistema, é importante que suas variáveis estejam organizadas por colunas, como no sistema (1.3).

A última coluna da matriz completa associada a um sistema é chamada de **coluna dos termos independentes**, e corresponde aos termos do lado direito das equações do sistema. Às vezes, convém enfatizar, graficamente, com uma linha vertical, a distinção entre a submatriz de coeficientes e a coluna dos termos independentes. Poderíamos escrever a matriz completa (1.3′) como

$$\left[\begin{array}{rrr|r} 1 & -4 & 1 & 0 \\ 0 & 3 & -6 & -3 \\ -3 & 14 & -6 & 2 \end{array}\right].$$

Na introdução do capítulo, mencionamos o procedimento de *escalonamento*, também chamado de *eliminação Gaussiana*. O objetivo desse procedimento é fornecer, para um dado sistema linear, um *sistema equivalente* que seja muito fácil de resolver.[2] Grosso modo, isso é realizado da seguinte forma: Usamos o termo em x_1 da primeira equação do sistema para eliminar os termos em x_1 das outras equações. Em seguida, usamos o termo em x_2 da segunda equação do sistema para eliminar os termos em x_2 das outras equações, e assim por diante, até a última variável do sistema. Veremos que o processo nem sempre funcionará exatamente desta forma, mas podemos ao menos encarar um primeiro exemplo usando esta ideia. O procedimento será descrito, em sua forma geral, na seção 1.5.

Vamos, então, achar o conjunto-solução do sistema (1.3) usando escalonamento. Para nos habituarmos com a notação matricial, vamos usá-la lado a lado com a notação usual neste primeiro exemplo:

$$\left\{\begin{array}{rcrcrcr} x_1 & - & 4x_2 & + & x_3 & = & 0 \\ & & 3x_2 & - & 6x_3 & = & -3 \\ -3x_1 & + & 14x_2 & - & 6x_3 & = & 2 \end{array}\right. \qquad \left[\begin{array}{rrr|r} 1 & -4 & 1 & 0 \\ 0 & 3 & -6 & -3 \\ -3 & 14 & -6 & 2 \end{array}\right] \tag{1.3″}$$

Vamos usar o termo em x_1 da primeira equação para eliminá-lo das outras. Como x_1 não aparece na segunda equação (o seu coeficiente já é zero), basta "zerar" seu coeficiente na terceira. Para isso, substituimos a terceira equação pela soma dela própria com três vezes a primeira equação:

$$\begin{array}{r} \text{equação 3 :} \\ +3\cdot(\text{equação 1}) : \\ \hline \text{nova equação 3 :} \end{array} \qquad \begin{array}{rrcrcrcr} & -3x_1 & + & 14x_2 & - & 6x_3 & = & 2 \\ +3(& x_1 & - & 4x_2 & + & x_3 & = & 0\ ) \\ \hline & & & 2x_2 & - & 3x_3 & = & 2 \end{array}$$

(Esta conta costuma ser feita "de cabeça", usando diretamente a forma matricial.) O novo sistema, portanto, é:

$$\left\{\begin{array}{rcrcrcr} x_1 & - & 4x_2 & + & x_3 & = & 0 \\ & & 3x_2 & - & 6x_3 & = & -3 \\ & & 2x_2 & - & 3x_3 & = & 2 \end{array}\right. \qquad \left[\begin{array}{rrr|r} 1 & -4 & 1 & 0 \\ 0 & 3 & -6 & -3 \\ 0 & 2 & -3 & 2 \end{array}\right] \tag{1.4}$$

Em seguida, usamos o termo em x_2 da segunda equação para eliminá-lo das outras. Mas, antes de fazê-lo, multiplicamos a segunda equação por $1/3$ para simplificar as contas:

$$\left\{\begin{array}{rcrcrcr} x_1 & - & 4x_2 & + & x_3 & = & 0 \\ & & x_2 & - & 2x_3 & = & -1 \\ & & 2x_2 & - & 3x_3 & = & 2 \end{array}\right. \qquad \left[\begin{array}{rrr|r} 1 & -4 & 1 & 0 \\ 0 & 1 & -2 & -1 \\ 0 & 2 & -3 & 2 \end{array}\right] \tag{1.5}$$

[2] Lembre-se de que sistemas equivalentes têm *o mesmo conjunto-solução*, conforme definimos ao final da seção anterior.

Agora, vamos eliminar o termo em x_2 da terceira equação:

$$\begin{array}{rr} \text{equação 3 :} & 2x_2 - 3x_3 = 2 \\ -2\cdot(\text{equação 2}) : & -2\big(x_2 - 2x_3 = -1\,\big) \\ \hline \text{nova equação 3 :} & x_3 = 4 \end{array}$$

O novo sistema é:

$$\begin{cases} x_1 - 4x_2 + x_3 = 0 \\ x_2 - 2x_3 = -1 \\ x_3 = 4 \end{cases} \qquad \left[\begin{array}{rrr|r} 1 & -4 & 1 & 0 \\ 0 & 1 & -2 & -1 \\ 0 & 0 & 1 & 4 \end{array}\right] \tag{1.6}$$

Desejamos, também, eliminar o termo em x_2 da primeira equação. Mas é mais eficiente eliminar, primeiro, os termos em x_3 da primeira e da segunda equação, usando a terceira (pois, assim, faremos menos contas):

$$\begin{array}{rr} \text{equação 1 :} & x_1 - 4x_2 + x_3 = 0 \\ -(\text{equação 3}) : & -\big(x_3 = 4\,\big) \\ \hline \text{nova equação 1 :} & x_1 - 4x_2 = -4 \end{array}$$

$$\begin{array}{rr} \text{equação 2 :} & x_2 - 2x_3 = -1 \\ +2\cdot(\text{equação 3}) : & +2\big(x_3 = 4\,\big) \\ \hline \text{nova equação 2 :} & x_2 = 7 \end{array}$$

Substituindo essas novas equações no sistema (1.6):

$$\begin{cases} x_1 - 4x_2 = -4 \\ x_2 = 7 \\ x_3 = 4 \end{cases} \qquad \left[\begin{array}{rrr|r} 1 & -4 & 0 & -4 \\ 0 & 1 & 0 & 7 \\ 0 & 0 & 1 & 4 \end{array}\right]$$

Agora sim, vamos eliminar o termo em x_2 da primeira equação, usando a segunda. Observe que, graças ao trabalho feito previamente com o x_3, não será necessário fazer mais conta alguma envolvendo os seus coeficientes:

$$\begin{array}{rr} \text{equação 1 :} & x_1 - 4x_2 = -4 \\ +4\cdot(\text{equação 2}) : & +4\big(x_2 = 7\,\big) \\ \hline \text{nova equação 1 :} & x_1 = 24 \end{array}$$

O novo sistema, finalmente, é:

$$\begin{cases} x_1 = 24 \\ x_2 = 7 \\ x_3 = 4 \end{cases} \qquad \left[\begin{array}{rrr|r} 1 & 0 & 0 & 24 \\ 0 & 1 & 0 & 7 \\ 0 & 0 & 1 & 4 \end{array}\right] \tag{1.7}$$

Este sistema é tão simples, que podemos imediatamente "ler" seu conjunto-solução: é o conjunto que contém, como único elemento, a lista $(x_1, x_2, x_3) = (24, 7, 4)$. Substitua esses valores no sistema original (1.3) para verificar que, de fato, encontramos uma solução. É importante conferir as suas contas *sempre*!

1.3 Operações-linha e equivalência de sistemas

Observe que, no exemplo da seção anterior, usamos apenas dois "truques" para progressivamente simplificar o sistema (1.3′′). Na passagem de (1.3′′) para (1.4), substituímos uma linha da matriz completa (a terceira) pela soma dela mesma a um múltiplo de outra linha (a primeira). Na passagem de (1.4) a (1.5), multiplicamos uma linha da matriz completa (a segunda) por uma constante. Usamos o primeiro "truque", nova e repetidamente, no trajeto de (1.5) até (1.7).

Os "truques" básicos que usaremos no processo de escalonamento, de fato, são *três*, conforme a definição a seguir.

Definição 1.1
As operações abaixo são chamadas de **operações elementares sobre as linhas** de uma matriz, ou, simplesmente, de **operações-linha**.

1. *Substituição*: substituir uma linha pela soma dela própria com um múltiplo de outra linha.
2. *Reescalonamento* ou *reescalamento*: multiplicar todos os elementos de uma linha por uma constante *diferente de zero*.
3. *Permutação*: permutar (ou seja, trocar, intercambiar) duas linhas entre si.

Cuidado!
As operações-linha são, como diz o nome, operações sobre *as linhas* de uma matriz. Quando estamos procurando o conjunto-solução de um sistema, *não faz sentido operar sobre as colunas* da matriz completa associada. Se o fizermos, chegaremos a um resultado incorreto. (A não ser que tenhamos *muita* sorte...)

É conveniente usar a seguinte notação para indicar operações-linha:

$\ell_i \to \ell_i + \alpha\ell_j$	Indica a *substituição* da i-ésima linha de uma matriz pela soma dela própria com α vezes a j-ésima linha, onde $j \neq i$.
$\ell_i \to \alpha\ell_i$	Indica o *reescalonamento* da i-ésima linha de uma matriz pelo fator α. Lembre que $\alpha \neq 0$.
$\ell_i \leftrightarrow \ell_j$	Indica a *permutação* das linhas i e j de uma matriz.

Com essa notação, não precisamos escrever coisas do tipo: "vamos permutar a segunda e a quinta linha". Podemos, simplesmente, indicar $\ell_2 \leftrightarrow \ell_5$. Enfatizamos que, na operação de reescalonamento, o fator α *não pode ser zero*, isto é, $\ell_i \to 0\ell_i$ *não* é uma operação-linha válida (a justificativa será dada na demonstração da proposição 1.2, logo adiante).

Para exemplificar o uso da notação acima, vamos reescrever dois passos do escalonamento da seção anterior, começando em (1.4):

$$\begin{bmatrix} 1 & -4 & 1 & 0 \\ 0 & 3 & -6 & -3 \\ 0 & 2 & -3 & 2 \end{bmatrix} \xrightarrow{\ell_2 \to \frac{1}{3}\ell_2} \begin{bmatrix} 1 & -4 & 1 & 0 \\ 0 & 1 & -2 & -1 \\ 0 & 2 & -3 & 2 \end{bmatrix} \xrightarrow{\ell_3 \to \ell_3 - 2\ell_2} \begin{bmatrix} 1 & -4 & 1 & 0 \\ 0 & 1 & -2 & -1 \\ 0 & 0 & 1 & 4 \end{bmatrix} \tag{1.8}$$

Proposição 1.2
As operações elementares sobre as linhas são reversíveis, isto é, para cada operação-linha, existe uma operação-linha reversa que a "desfaz".

Demonstração: A operação $\ell_i \to \ell_i + \alpha\ell_j$ é "desfeita" pela operação $\ell_i \to \ell_i - \alpha\ell_j$ (verifique!). A operação $\ell_i \to \alpha\ell_i$ é revertida por $\ell_i \to \frac{1}{\alpha}\ell_i$. (Foi por isso, a propósito, que exigimos a condição $\alpha \neq 0$ na operação de reescalonamento. Do contrário, não seria possível revertê-la!) Finalmente, a operação $\ell_i \leftrightarrow \ell_j$ é revertida aplicando-se, novamente, ela própria. Note que trocar ℓ_i por ℓ_j e, em seguida, ℓ_j por ℓ_i é o mesmo que não fazer coisa alguma. $\square$

Uma *sequência* de operações-linha também pode ser revertida, operação por operação. A sequência das duas operações em (1.8), por exemplo, é revertida por meio da operação $\ell_3 \to \ell_3 + 2\ell_2$ seguida de $\ell_2 \to 3\ell_2$. Verifique isto, observando que é necessário atentar para a *ordem* em que as operações são realizadas.

Definição 1.3
Dizemos que duas matrizes são **equivalentes por operações sobre as linhas**, ou simplesmente **linha-equivalentes**, se existe uma sequência de operações-linha que leva uma matriz à outra.

Em vista da proposição 1.2, dizer que uma matriz A é linha-equivalente a uma matriz B é o mesmo que dizer que B é linha-equivalente a A. As três matrizes que aparecem em (1.8), por exemplo, são todas linha-equivalentes.

O teorema a seguir é importante porque dá embasamento à estratégia de simplificar um sistema aplicando operações-linha em sua matriz completa. Temos a garantia de que a versão simplificada do sistema terá o mesmo conjunto-solução que o sistema original.

Teorema 1.4
Se dois sistemas lineares têm matrizes completas que são linha-equivalentes, então os dois sistemas são equivalentes, isto é, têm o mesmo conjunto-solução.

A prova do teorema é dada abaixo. *O aluno sem interesse em argumentos matemáticos pode passar diretamente à seção 1.4.*

Demonstração: Suponha que A e B sejam matrizes associadas a sistemas lineares — vamos chamá-los de "sistema (A)" e "sistema (B)" — e que a matriz B seja obtida de A via uma operação-linha qualquer. Afirmamos que os sistemas (A) e (B) têm o mesmo conjunto-solução, isto é, qualquer solução do sistema (A) é também uma solução do sistema (B), e *vice-versa*.

De fato, suponha que $A \xrightarrow{\ell_i \to \ell_i + \alpha\ell_j} B$. Se a lista $\mathbf{s} = (s_1, s_2, \ldots, s_n)$ é uma solução do sistema (A), então $\mathbf{s}$ satisfaz a todas as equações deste sistema, em particular a i-ésima e a j-ésima. Estas equações estão associadas, respectivamente, às linhas ℓ_i e ℓ_j da matriz A. Portanto, $\mathbf{s}$ também satisfaz a equação associada a $\ell_i + \alpha\ell_j$, isto é, a lista $\mathbf{s}$ satisfaz a i-ésima equação do sistema (B). (Verifique esse passo do argumento! Considere exemplos, se preferir.) Todas as outras equações do sistema (B) são idênticas às equações correspondentes de (A), pois só "mexemos" na i-ésima equação. Portanto, $\mathbf{s}$ também satisfaz essas outras equações. Logo, a lista $\mathbf{s}$ é uma solução do sistema (B). A parte do "vice-versa" agora é fácil. Basta usar a reversibilidade das operações-linha: como $A \xrightarrow{\ell_i \to \ell_i + \alpha\ell_j} B$, vale também que $B \xrightarrow{\ell_i \to \ell_i - \alpha\ell_j} A$. Portanto, podemos repetir o argumento acima para mostrar que qualquer solução do sistema (B) é também uma solução de (A). Verifique que o argumento deste parágrafo pode ser adaptado às operações de reescalamento e de permutação.

Para terminar a prova do teorema, recordemos a seguinte definição: duas matrizes são linha-equivalentes se existe uma sequência de operações-linha que leva uma à outra. Aplicando, sucessivamente, o argumento acima em cada um dos passos da sequência, obtemos o resultado desejado. Todas as matrizes envolvidas, da primeira à última, representam sistemas com o mesmo conjunto-solução. $\square$

1.4 Formas escalonadas

Considere o sistema linear abaixo, apresentado com sua matriz completa:

$$\begin{cases} x_1 - 5x_2 + 2x_3 - 4x_4 = 0 \\ \qquad\qquad\; - 3x_3 + \;\; x_4 = 9 \\ \qquad\qquad\qquad\qquad 2x_4 = 6 \end{cases} \qquad \left[\begin{array}{cccc|c} 1 & -5 & 2 & -4 & 0 \\ 0 & 0 & -3 & 1 & 9 \\ 0 & 0 & 0 & 2 & 6 \end{array}\right] \tag{1.9}$$

É fácil encontrar o conjunto-solução desse sistema. A última equação implica, imediatamente, que $x_4 = 3$. A segunda equação, portanto, se reduz a

$$-3x_3 + 3 = 9 \qquad \text{ou a} \qquad -3x_3 = 6 \qquad \text{ou, ainda, a} \qquad x_3 = -2.$$

Agora, podemos colocar $x_3 = -2$ e $x_4 = 3$ na primeira equação, obtendo

$$x_1 - 5x_2 + 2 \cdot (-2) - 4 \cdot 3 = 0 \qquad \text{ou} \qquad x_1 = 16 + 5x_2.$$

O conjunto-solução do sistema (1.9), portanto, é descrito por

$$\begin{cases} x_1 = 16 + 5x_2, \\ x_2 \text{ é uma "variável livre" (veja o comentário abaixo),} \\ x_3 = -2, \\ x_4 = 3. \end{cases} \tag{1.10}$$

A solução do sistema não é única, isto é, o conjunto-solução tem mais do que um elemento. De fato, podemos arbitrar ("chutar") qualquer valor para a "variável livre" x_2. Para cada valor arbitrado, obtemos uma solução diferente. Portanto, há uma infinidade de soluções. Nas seções 1.7 e 1.8, este assunto será aprofundado. O significado de "variável livre" será dado mais precisamente na subseção 1.7.1.

Ao invés de resolver o sistema (1.9) usando substituições, como fizemos acima, poderíamos ter utilizado operações-linha para simplificá-lo ainda mais:

$$\left[\begin{array}{cccc|c} 1 & -5 & 2 & -4 & 0 \\ 0 & 0 & -3 & 1 & 9 \\ 0 & 0 & 0 & 2 & 6 \end{array}\right] \xrightarrow{\ell_3 \to \frac{1}{2}\ell_3} \left[\begin{array}{cccc|c} 1 & -5 & 2 & -4 & 0 \\ 0 & 0 & -3 & 1 & 9 \\ 0 & 0 & 0 & 1 & 3 \end{array}\right] \xrightarrow[\ell_2 \to \ell_2 - \ell_3]{\ell_1 \to \ell_1 + 4\ell_3}$$
$$\longrightarrow \left[\begin{array}{cccc|c} 1 & -5 & 2 & 0 & 12 \\ 0 & 0 & -3 & 0 & 6 \\ 0 & 0 & 0 & 1 & 3 \end{array}\right] \xrightarrow{\ell_2 \to -\frac{1}{3}\ell_2} \left[\begin{array}{cccc|c} 1 & -5 & 2 & 0 & 12 \\ 0 & 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 1 & 3 \end{array}\right] \xrightarrow{\ell_1 \to \ell_1 - 2\ell_2}$$
$$\longrightarrow \left[\begin{array}{cccc|c} 1 & -5 & 0 & 0 & 16 \\ 0 & 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 1 & 3 \end{array}\right] \tag{1.11}$$

A última matriz obtida acima representa um sistema tão simples que, assim como no caso do sistema (1.7), podemos “ler” a descrição (1.10) de seu conjunto-solução.[3]

Podemos formular, agora, a pergunta-chave desta seção. Os sistemas associados às matrizes que aparecem em (1.7), (1.9) e (1.11) são muito simples de resolver. Nesse sentido, *qual é a particularidade dessas matrizes?*

A resposta está na definição abaixo. Antes de apresentá-la, precisamos de alguns termos auxiliares. Dizemos que uma linha, ou uma coluna, de uma matriz é **não-nula** se ela contém, ao menos, um elemento diferente de zero. O **elemento líder** de uma linha é o primeiro elemento diferente de zero desta linha, olhando da esquerda para a direita. Uma linha só de zeros não tem elemento líder.

Definição 1.5
Dizemos que uma matriz está em **forma escalonada**, ou, simplesmente, que é uma **matriz escalonada**, se ela tem as duas propriedades abaixo:

1. As linhas não-nulas estão todas acima de qualquer linha só de zeros. Ou seja, se há linhas só de zeros, elas estão todas agrupadas na parte de baixo da matriz.

2. O elemento líder de cada linha não-nula está numa coluna à direita do elemento líder da linha acima.

Dizemos que uma matriz está na **forma escalonada reduzida**, ou que é uma **matriz escalonada reduzida**, se ela tem as propriedades seguintes, além das anteriores:

3. O elemento líder de cada linha não-nula é igual a 1.

4. Cada elemento líder é o único elemento diferente de zero da sua coluna.

[3] Mas essa “leitura” requer um pouco de prática! Recomendamos que o leitor invista um ou dois minutos de atenção, agora, para perceber como se pode obter (1.10) a partir de (1.11).

A propriedade 2 diz que os elementos líderes formam uma escada que desce da esquerda para a direita na matriz — é por isso que usamos o adjetivo *escalonado*, que significa "ter forma de escada". Uma consequência simples dessa propriedade é que, em uma matriz escalonada, *todos os elementos que estão abaixo de um elemento líder são zeros*.

A matriz completa do sistema (1.9), por exemplo, está em forma escalonada, mas *não* está na forma escalonada reduzida:

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

Hachuramos cada linha a partir dos elementos líderes, em destaque, a fim de realçar a "forma de escada" da matriz. Observe que a matriz acima, de fato, tem a propriedade 2: O elemento líder $\boxed{2}$ da linha 3 está numa coluna à direita do elemento líder $\boxed{-3}$ da linha anterior. Este $\boxed{-3}$, por sua vez, também está numa coluna à direita do elemento líder $\boxed{1}$ da linha acima (*não* é necessário que ele esteja na coluna *imediatamente* à direita). Como não há linhas só de zeros, não há o que verificar quanto à propriedade 1.

A última matriz que obtivemos em (1.11), por sua vez, não só está em forma escalonada, como também está na forma *reduzida*:

$$\begin{bmatrix} \boxed{1} & -5 & 0 & 0 & 16 \\ 0 & 0 & \boxed{1} & 0 & -2 \\ 0 & 0 & 0 & \boxed{1} & 3 \end{bmatrix}$$

Observe que os elementos líderes, novamente em destaque, são todos iguais a um, e são as únicas entradas não-nulas em suas respectivas colunas. A matriz acima, portanto, tem também as propriedades 3 e 4.

Agora, vejamos exemplos de matrizes que *não* estejam em forma escalonada:

Uma atualização do "pacote" `colortbl` "estragou" a formatação dessas matrizes. O problema será corrigido oportunamente...

$$\begin{bmatrix} \boxed{1} & -3 & 1 & -5 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \boxed{9} & -8 & 1 & 4 \\ 0 & 0 & 0 & \boxed{3} & 6 \end{bmatrix} \qquad \begin{bmatrix} \boxed{1} & -3 & 1 & -5 & 0 \\ 0 & \boxed{5} & -2 & 0 & 7 \\ 0 & \boxed{9} & -8 & 1 & 4 \\ 0 & 0 & 0 & \boxed{3} & 6 \end{bmatrix}$$

A primeira matriz acima não é uma matriz escalonada, já que viola a propriedade 1: há uma linha só de zeros no meio da matriz, ou seja, há uma linha não-nula abaixo de uma linha de zeros. A segunda matriz não está em forma escalonada, já que viola a propriedade 2: o elemento líder $\boxed{9}$ da terceira linha não está em uma coluna à direita do elemento líder $\boxed{5}$ da linha acima. Esta segunda matriz, apesar de aparentar uma forma de escada, tem um "degrau" grande demais na segunda coluna.

A seguinte notação será útil para indicar formas gerais de matrizes escalonadas:

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

As entradas $\blacksquare$ são os elementos líderes, que podem ter qualquer valor *diferente de zero*. As entradas indicadas por $*$ podem ter qualquer valor (zero ou não). Ambas as matrizes "genéricas" acima têm as propriedades 1 e 2 (verifique!), portanto são, de fato, escalonadas.

Idem: a formatação dessas matrizes será corrigida oportunamente.

As matrizes a seguir são formas gerais de matrizes escalonadas *reduzidas*:

$$\begin{bmatrix} \boxed{1} & * & 0 & * & * & * & * \\ 0 & 0 & \boxed{1} & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & \boxed{1} & 0 & * & * & * & 0 & * & 0 \\ 0 & 0 & \boxed{1} & * & * & * & 0 & * & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & * & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{1.12}$$

Verifique que estas matrizes têm as quatro propriedades da definição 1.5.

Se B for uma matriz escalonada linha-equivalente a A, diremos que B é **uma forma escalonada da matriz** A. Se B estiver na forma escalonada reduzida, diremos que B é **a forma escalonada reduzida da matriz** A. O processo que leva uma matriz, por operações-linha, até uma de suas formas escalonadas é chamado de **escalonamento**, como já mencionamos.

Em geral, uma matriz tem muitas formas escalonadas. Todas as matrizes que aparecem em (1.11) são formas escalonadas distintas de uma mesma matriz, por exemplo. No entanto, cada matriz tem apenas uma forma escalonada reduzida, isto é, *cada matriz é linha-equivalente a uma única matriz escalonada reduzida.* É por isso que usamos as expressões "*uma* forma escalonada" e "*a* forma escalonada reduzida" de uma matriz.

Um sistema representado por uma matriz escalonada é simples de analisar e de resolver, especialmente se a matriz for escalonada reduzida, como você poderá observar em vários exemplos e exercícios — a começar pelos sistemas (1.7) e (1.9). A definição 1.5 responde, portanto, à "pergunta-chave" da seção. Os conceitos e a terminologia que introduzimos aqui, no entanto, aplicam-se a *qualquer matriz*, e não apenas a matrizes completas associadas a sistemas lineares.

Observação

É comum dizermos coisas como "a matriz está em forma escalonada" e "B é uma forma escalonada de A". Apesar de consagrada, esta terminologia pode induzir a um erro conceitual, que queremos prevenir aqui. Parece-nos mais correto dizer "a matriz *é* escalonada" ou "a matriz *tem* forma de escada" que dizer "a matriz *está* escalonada", visto que o verbo "estar" sugere uma transitoriedade que é falsa. O fato é que uma matriz A ou é escalonada ou não é. Se A não for uma matriz escalonada, ela não poderá "tornar-se" escalonada nunca. O processo de escalonamento produz *outra* matriz B, que é escalonada e linha-equivalente a A. Mas o processo não faz com que a própria matriz A "vire" escalonada.

1.5 O algoritmo de escalonamento

Começamos a seção com uma breve anedota (não muito boa, admitimos). Uma turista carioca chega à estação das barcas em Niterói e pergunta a um transeunte como se chega ao Teatro Municipal. Infelizmente, esse transeunte é professor do Instituto de Matemática da UFF, e sua resposta é absolutamente correta, porém absolutamente inútil: "Ah, é fácil! Você chega lá movendo, alternadamente, suas pernas direita e esquerda."

A desventurada turista sabe *aonde* quer chegar (ao Teatro Municipal de Niterói), sabe quais são os *passos admissíveis* para chegar lá (são, literalmente, passos, ou, nas palavras do matemático, "movimentos alternados das pernas"), mas não faz ideia da direção dos passos a tomar.

Encontramo-nos em uma situação curiosa, análoga à da turista. Dada uma matriz qualquer, sabemos *aonde* queremos chegar (a uma forma escalonada linha-equivalente) e sabemos quais são os *passos admissíveis* a tomar (são as operações elementares sobre as linhas). Mas ainda não discutimos o trajeto para se chegar lá. Quais são, exatamente, as operações-linha que devemos aplicar, e em que ordem devemos aplicá-las?

Nesta seção, abordaremos o *algoritmo de escalonamento*, que responde à questão acima. Um **algoritmo** é um procedimento (ou uma "receita") para resolver um determinado problema matemático ou computacional, e que, usualmente, é descrito passo a passo. Frequentemente, o mesmo passo tem que ser aplicado diversas vezes em diferentes etapas do processo, e o algoritmo de escalonamento não é exceção.

Antes de começarmos, precisamos explicar o significado de dois termos que serão empregados. Quando usamos um elemento de uma matriz para "zerar" as entradas que estão abaixo ou acima, dizemos que esse elemento é o *pivô* das operações-linha envolvidas. Uma coluna que contém um pivô é chamada de *coluna-pivô*. Este último conceito será definido mais precisamente na seção 1.6.

1.5.1 Obtenção de uma forma escalonada

Os passos do algoritmo de escalonamento são dados a seguir. Cada um é ilustrado por meio do seguinte exemplo: achar uma forma escalonada da matriz

$$A = \begin{bmatrix} 0 & -3 & -6 & 4 & 9 \\ -1 & -2 & -1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ 1 & 4 & 5 & -9 & -7 \end{bmatrix}. \tag{1.13}$$

Passo 1: Olhando a matriz da esquerda para a direita, procure a primeira coluna não-nula. Esta será a coluna-pivô do próximo passo.

A coluna 1 da matriz A é não-nula (não é só de zeros), então esta será nossa primeira coluna-pivô.

Passo 2: Escolha qualquer elemento não-nulo da coluna-pivô para servir de pivô. Você pode escolher um que torne fáceis as contas que estão por vir. Com prática, você será capaz de antever as contas e aplicar este critério.

Na primeira coluna de A, que é a coluna-pivô no momento, podemos escolher o 1 que está na última linha para ser o pivô. Vamos destacá-lo na matriz:

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

A escolha deste pivô é conveniente. Se tivéssemos escolhido o -2 na terceira linha, por exemplo, os cálculos adiante envolveriam frações.

Passo 3: Se for necessário, passe o pivô para a primeira linha disponível, usando uma operação de permutação ($\ell_i \leftrightarrow \ell_j$).

Nosso pivô está na quarta linha da matriz, portanto temos que trocar as linhas 1 e 4:

$$\begin{bmatrix} 0 & -3 & -6 & 4 & 9 \\ -1 & -2 & -1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ \boxed{1} & 4 & 5 & -9 & -7 \end{bmatrix} \xrightarrow{\ell_1 \leftrightarrow \ell_4} \begin{bmatrix} \boxed{1} & 4 & 5 & -9 & -7 \\ -1 & -2 & -1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix}$$

Se o pivô já estivesse na linha 1, a troca seria desnecessária aqui.

Passo 4: Opcionalmente, use operações de reescalonamento ($\ell_i \to \alpha\ell_i$) sobre quaisquer linhas que você queira, para facilitar as contas do próximo passo. Mas lembre-se de que $\alpha \neq 0$!

O passo 4 é desnecessário neste estágio do nosso exemplo.

Passo 5: Use operações de substituição ($\ell_i \to \ell_i + \alpha\ell_j$) para "zerar" todos os elementos da coluna-pivô que estão abaixo do pivô.

Em nosso exemplo, temos que "zerar" o -1 e o -2 nas linhas 2 e 3, respectivamente:

$$\begin{bmatrix} \boxed{1} & 4 & 5 & -9 & -7 \\ -1 & -2 & -1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix} \xrightarrow[\ell_3 \to \ell_3 + 2\ell_1]{\ell_2 \to \ell_2 + \ell_1} \begin{bmatrix} \boxed{1} & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 5 & 10 & -15 & -15 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix}$$

O trabalho com o primeiro pivô terminou. A ideia, agora, é repetir todos os passos acima, mas olhando apenas para a parte da matriz que está *abaixo* da linha do pivô. Formalizamos isto no passo seguinte.

Passo 6: Ignore (cubra com a mão, se quiser) a linha que contém o pivô, e todas as linhas acima dela (se houver alguma). Aplique os passos de 1 a 5 sobre a submatriz restante. Continue repetindo este procedimento todo até chegar a uma forma escalonada, ou seja, até que não haja linhas restantes, ou que estas sejam todas linhas de zeros.

Em nosso exemplo, temos que "cobrir", por enquanto, apenas a linha 1, que contém o pivô (não há outras linhas acima dela). Abaixo, as linhas "cobertas" (ou ignoradas) serão impressas em tom mais claro:

$$\begin{bmatrix} \boxed{1} & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 5 & 10 & -15 & -15 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix}$$

Agora, voltamos ao passo 1: procurar a primeira coluna não-nula na submatriz indicada acima (esta será a próxima coluna-pivô). Cuidado: A primeira coluna não-nula é a coluna 2! O único elemento não-nulo da coluna 1 é o $\boxed{1}$, o pivô anterior. Mas esse elemento está em uma linha ignorada. Então, a coluna-pivô "da vez" é a coluna 2:

$$\begin{bmatrix} \boxed{1} & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 5 & 10 & -15 & -15 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix}$$

Passo 2: Escolher um elemento não-nulo na coluna-pivô. Podemos escolher qualquer entrada na segunda coluna, exceto o 4 da primeira linha, pois esta é uma linha ignorada. Escolhemos o 2, que destacamos abaixo, para ser pivô:

$$\begin{bmatrix} \boxed{1} & 4 & 5 & -9 & -7 \\ 0 & \boxed{2} & 4 & -6 & -6 \\ 0 & 5 & 10 & -15 & -15 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix}$$

Passo 3: Se necessário, passar o pivô (por troca de linhas) para a primeira linha disponível. Neste caso, isso não será necessário, pois o pivô *já está* na primeira linha disponível. Como a linha 1 é ignorada, ela não está disponível.

Passo 4: Usar operações de reescalonamento, opcionalmente, para facilitar as contas posteriores. Em nosso exemplo, é conveniente dividir a segunda linha por 2 e a terceira linha por 5:

$$\begin{bmatrix} \boxed{1} & 4 & 5 & -9 & -7 \\ 0 & \boxed{2} & 4 & -6 & -6 \\ 0 & 5 & 10 & -15 & -15 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix} \xrightarrow[\ell_3 \to \frac{1}{5}\ell_3]{\ell_2 \to \frac{1}{2}\ell_2} \begin{bmatrix} \boxed{1} & 4 & 5 & -9 & -7 \\ 0 & \boxed{1} & 2 & -3 & -3 \\ 0 & 1 & 2 & -3 & -3 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix}$$

Passo 5: “Zerar” os elementos abaixo do pivô:

$$\begin{bmatrix} \boxed{1} & 4 & 5 & -9 & -7 \\ 0 & \boxed{1} & 2 & -3 & -3 \\ 0 & 1 & 2 & -3 & -3 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix} \xrightarrow[\ell_4 \to \ell_4 + 3\ell_2]{\ell_3 \to \ell_3 - \ell_2} \begin{bmatrix} \boxed{1} & 4 & 5 & -9 & -7 \\ 0 & \boxed{1} & 2 & -3 & -3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -5 & 0 \end{bmatrix}$$

Chegamos, finalmente, ao passo 6: “Cobrir” a linha que contém o pivô e as linhas acima dela, e aplicar o processo todo, novamente, à submatriz restante. Em nosso exemplo, temos que cobrir a linha 2, que contém o pivô “da vez”. A linha 1, que está acima dela, permanece coberta também:

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

Agora, retornamos novamente ao passo 1! As três primeiras colunas são só de zeros, então a nova coluna-pivô será a quarta:

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

Passos 2 e 3: O novo pivô é, necessariamente, o -5 na coluna destacada acima (um elemento zero nunca pode ser pivô). Temos, então, que passar o pivô para a primeira linha disponível, que é a linha 3 (as linhas 1 e 2 estão “indisponíveis”):

$$\begin{bmatrix} \boxed{1} & 4 & 5 & -9 & -7 \\ 0 & \boxed{1} & 2 & -3 & -3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \boxed{-5} & 0 \end{bmatrix} \xrightarrow{\ell_3 \leftrightarrow \ell_4} \begin{bmatrix} \boxed{1} & 4 & 5 & -9 & -7 \\ 0 & \boxed{1} & 2 & -3 & -3 \\ 0 & 0 & 0 & \boxed{-5} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

O processo terminou, porque chegamos a uma matriz escalonada. De fato, o passo 4 é opcional, e o passo 5 é desnecessário, pois já temos um zero no único elemento abaixo do pivô $\boxed{-5}$. Ao aplicar o passo 6, cobrimos a terceira linha (a do pivô). O que sobra é uma linha só de zeros. Dessa forma, obtivemos a matriz escalonada

$$C = \begin{bmatrix} \boxed{1} & 4 & 5 & -9 & -7 \\ 0 & \boxed{1} & 2 & -3 & -3 \\ 0 & 0 & 0 & \boxed{-5} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \text{cuja forma geral é} \quad \begin{bmatrix} \blacksquare & * & * & * & * \\ 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & \blacksquare & * \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{1.14}$$

Perceba que os pivôs tornam-se os elementos líderes na forma escalonada. Deixamos os pivôs em destaque, pois eles serão usados, novamente, abaixo.

1.5.2 Obtenção da forma escalonada reduzida

A matriz C, que obtivemos acima, é linha-equivalente à matriz A de (1.13). Ela está em forma escalonada, mas não está na forma escalonada *reduzida*. Veremos, adiante, que podemos extrair muita informação sobre uma matriz analisando qualquer uma de suas formas escalonadas (nem sempre é necessário chegar à forma reduzida).

Às vezes, no entanto, é necessário obter a forma reduzida. É fácil, por exemplo, "ler" o conjunto-solução de um sistema linear a partir da forma escalonada reduzida de sua matriz completa (como vimos na seção 1.4). Mas não é tão fácil fazer essa "leitura" a partir de uma forma escalonada qualquer.

Para obter a forma escalonada reduzida de uma matriz, primeiro, encontramos uma forma escalonada qualquer (para isso, aplicamos os passos descritos acima). Em seguida, aplicamos o seguinte passo adicional:

Passo 7: Use operações de substituição ($\ell_i \to \ell_i + \alpha\ell_j$) para "zerar" todos os elementos *acima* de cada pivô. Recomendamos que você comece com o pivô mais à direita da matriz, e prossiga, de pivô em pivô, da direita para a esquerda. Proceder nesta ordem reduz a quantidade total de cálculos. Para cada pivô que for diferente de 1, use uma operação de reescalonamento em sua linha para que ele se torne 1.

Se desejar, use operações de reescalonamento de linhas ($\ell_i \to \alpha\ell_i$), antes das operações de substituição, para facilitar as contas. Mas lembre-se de reescalonar as linhas no final para que todos os pivôs fiquem iguais a 1.

Vamos aplicar o passo 7 à matriz escalonada C que obtivemos em (1.14). Começamos com o pivô mais à direita, como sugerido, que é o $\boxed{-5}$ da quarta coluna. Para facilitar as contas, vamos reescalonar a linha do pivô para que ele se torne $\boxed{1}$ (de toda a forma, seria necessário fazer isso mais adiante):

$$\begin{bmatrix} \boxed{1} & 4 & 5 & -9 & -7 \\ 0 & \boxed{1} & 2 & -3 & -3 \\ 0 & 0 & 0 & \boxed{-5} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{\ell_3 \to -\frac{1}{5}\ell_3} \begin{bmatrix} \boxed{1} & 4 & 5 & -9 & -7 \\ 0 & \boxed{1} & 2 & -3 & -3 \\ 0 & 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Agora, vamos zerar as entradas acima do pivô "da vez". Indicamos o "pivô da vez" pela caixa preta. Os outros pivôs tem caixa mais clara.

$$\begin{bmatrix} \boxed{1} & 4 & 5 & -9 & -7 \\ 0 & \boxed{1} & 2 & -3 & -3 \\ 0 & 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow[\ell_2 \to \ell_2 + 3\ell_3]{\ell_1 \to \ell_1 + 9\ell_3} \begin{bmatrix} \boxed{1} & 4 & 5 & 0 & -7 \\ 0 & \boxed{1} & 2 & 0 & -3 \\ 0 & 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

O próximo pivô, indo da direita para a esquerda, é o $\boxed{1}$ na segunda coluna. Zeramos a entrada acima dele:

$$\begin{bmatrix} \boxed{1} & 4 & 5 & 0 & -7 \\ 0 & \boxed{1} & 2 & 0 & -3 \\ 0 & 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{\ell_1 \to \ell_1 - 4\ell_2} \begin{bmatrix} \boxed{1} & 0 & -3 & 0 & 5 \\ 0 & \boxed{1} & 2 & 0 & -3 \\ 0 & 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Se tivéssemos começado com este pivô, neste passo teríamos que fazer contas na quarta coluna. É por isso que é melhor seguir da direita para a esquerda.

O próximo pivô, à esquerda, é o $\boxed{1}$ da primeira coluna. Mas não há elementos acima dele para zerar! Como todos os pivôs já são iguais a 1, já obtivemos a forma escalonada reduzida da matriz A de (1.13):

$$B = \begin{bmatrix} \boxed{1} & 0 & -3 & 0 & 5 \\ 0 & \boxed{1} & 2 & 0 & -3 \\ 0 & 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{1.15}$$

1.6 Posições e colunas-pivô

Compare a matriz C de (1.14) com a B de (1.15), e note que as posições dos elementos líderes de cada linha não mudaram. Não é difícil perceber que as posições dos líderes nunca mudam quando passamos por operações-linha de uma forma escalonada qualquer para a forma escalonada reduzida. Como mencionamos antes, cada matriz tem uma única forma escalonada reduzida.

Dessa discussão, segue o seguinte fato fundamental: *As posições dos elementos líderes são sempre as mesmas em qualquer uma das formas escalonadas de uma dada matriz.* Essas posições são chamadas de **posições-pivô** da matriz, pois os elementos nessas posições são usados como pivôs no processo de escalonamento.

Uma coluna que contém uma posição-pivô em uma matriz é chamada de **coluna-pivô**, como você deve ter observado na seção anterior. As demais colunas são chamadas de **colunas não-pivô**.

Exemplo 1.6
Localize as posições-pivô da matriz A de (1.13) (página 13). Quais são as colunas-pivô de A? Quais são as colunas não-pivô?

Solução: Temos que determinar as posições dos elementos líderes em uma forma escalonada (qualquer) de A. Na seção 1.5.1, já encontramos uma forma escalonada de A, a saber, a matriz C de (1.14), que repetimos abaixo:

$$C = \begin{bmatrix} \boxed{1} & 4 & 5 & -9 & -7 \\ 0 & \boxed{1} & 2 & -3 & -3 \\ 0 & 0 & 0 & \boxed{-5} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \text{cuja forma geral é} \quad \begin{bmatrix} \blacksquare & * & * & * & * \\ 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & \blacksquare & * \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Os elementos líderes de C estão destacados, e suas posições são aquelas assinaladas por $\blacksquare$ na "forma geral". Estas são as posições-pivô de A, destacadas novamente abaixo, na própria matriz A. As colunas-pivô de A são, portanto, a primeira, a segunda, e a quarta (hachuradas abaixo); a terceira e a quinta são colunas não-pivô.

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

□

É importante entender que os conceitos de posição-pivô e coluna-pivô são definidos para *qualquer matriz*, e não apenas para matrizes em forma escalonada. Apesar de *não* ser uma matriz escalonada, a matriz A do exemplo acima tem posições-pivô bem definidas — só que foi necessário escalonar para "desvendá-las". As posições-pivô de uma matriz escalonada, por outro lado, são evidentes: são as posições dos elementos líderes.

Observe, também, que *não* é necessário obter a forma escalonada *reduzida* de uma matriz para determinar suas posições-pivô: basta encontrar uma forma escalonada *qualquer*. No exemplo acima, usamos a matriz escalonada C, que não é reduzida, para localizar as posições-pivô de A.

As posições e colunas-pivô têm papel central na análise de diversas questões de álgebra linear, como veremos nas seções a seguir e nos próximos capítulos.

1.7 Resolução de sistemas lineares

Nesta seção, vamos voltar ao tema que motivou todo este capítulo: a resolução de sistemas lineares. **Resolver** um sistema significa obter uma descrição de seu conjunto-solução, ou determinar que o sistema não tem solução alguma. Vamos começar com dois exemplos que vão nos auxiliar na abordagem de novos conceitos e que servem, também, para recapitular os já estudados.

Exemplo 1.7
Vamos resolver o seguinte sistema:

$$\begin{cases} x_1 + 2x_2 + 2x_3 - x_4 = 17 \\ -x_1 - 2x_2 + x_3 - 8x_4 = 4 \\ 2x_1 + 4x_2 + x_3 + 7x_4 = 13 \end{cases} \tag{1.16}$$

Atacamos este problema usando o método de escalonamento desenvolvido neste capítulo. Incialmente, escrevemos a matriz completa associada ao sistema:

$$G = \left[\begin{array}{rrrr|r} 1 & 2 & 2 & -1 & 17 \\ -1 & -2 & 1 & -8 & 4 \\ 2 & 4 & 1 & 7 & 13 \end{array}\right].$$

Vamos, agora, escalonar a matriz completa G:

$$\begin{bmatrix} \boxed{1} & 2 & 2 & -1 & 17 \\ -1 & -2 & 1 & -8 & 4 \\ 2 & 4 & 1 & 7 & 13 \end{bmatrix} \xrightarrow[\ell_3 \to \ell_3 - 2\ell_1]{\ell_2 \to \ell_2 + \ell_1} \begin{bmatrix} \boxed{1} & 2 & 2 & -1 & 17 \\ 0 & 0 & \boxed{3} & -9 & 21 \\ 0 & 0 & -3 & 9 & -21 \end{bmatrix} \longrightarrow$$

$$\xrightarrow{\ell_3 \to \ell_3 + \ell_2} \begin{bmatrix} \boxed{1} & 2 & 2 & -1 & 17 \\ 0 & 0 & \boxed{3} & -9 & 21 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Passamos à forma escalonada *reduzida*:

$$\begin{bmatrix} \boxed{1} & 2 & 2 & -1 & 17 \\ 0 & 0 & \boxed{3} & -9 & 21 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{\ell_2 \to \frac{1}{3}\ell_2} \begin{bmatrix} \boxed{1} & 2 & 2 & -1 & 17 \\ 0 & 0 & \boxed{1} & -3 & 7 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \longrightarrow$$

$$\xrightarrow{\ell_1 \to \ell_1 - 2\ell_2} H = \left[\begin{array}{cccc|c} \boxed{1} & 2 & 0 & 5 & 3 \\ 0 & 0 & \boxed{1} & -3 & 7 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Esta última matriz, que chamamos de H, representa o sistema

$$\begin{cases} x_1 + 2x_2 \qquad + 5x_4 = 3 \\ \qquad\qquad x_3 - 3x_4 = 7 \\ \qquad\qquad\qquad 0 \;= 0. \end{cases} \tag{1.17}$$

Por ser toda de zeros, a última linha de H representa a equação $0 = 0$, que podemos descartar sem perder informação alguma.

Como as matrizes G e H são linha-equivalentes (H foi obtida de G por operações-linha), *o sistema* (1.16) *é equivalente ao sistema* (1.17)*, isto é, esses dois sistemas têm o mesmo conjunto-solução.* Usamos, aqui, o teorema 1.4 da página 8.

Nosso problema se reduziu, portanto, a achar o conjunto-solução de (1.17). Mas este sistema é muito simples, visto que a matriz H que o representa está na forma escalonada reduzida.

Observe que as variáveis x_1 e x_3 correspondem a colunas-pivô da matriz H, ou da matriz G, enquanto x_2 e x_4 correspondem a colunas não-pivô. É fácil constatar estes fatos simplesmente olhando para a matriz H. É usual, neste caso, estabelecer x_1 e x_3 como *variáveis básicas* (ou *dependentes*); e x_2 e x_4, como *variáveis livres*.[4] Escrevendo as variáveis dependentes x_1 e x_3, em termos das variáveis livres x_2 e x_4, obtemos

$$\begin{cases} x_1 = 3 - 2x_2 - 5x_4, \\ x_2 \text{ é uma variável livre}, \\ x_3 = 7 + 3x_4, \\ x_4 \text{ é uma variável livre}. \end{cases} \tag{1.18}$$

Com um pouco de prática, fica fácil "ler" (1.18) diretamente da matriz H, sem a necessidade de escrever o sistema (1.17).

Concluímos que (1.18) descreve o conjunto-solução do sistema (1.16). Esse sistema tem uma infinidade de soluções: uma para cada par de valores arbitrado para as variáveis livres x_2 e x_3. Já havíamos nos deparado com um caso semelhante: reveja o sistema (1.9), na página 9, e seu conjunto-solução descrito em (1.10).

[4]Esses conceitos serão esclarecidos na subseção a seguir.

Exemplo 1.8
Consideramos o sistema:

$$\begin{cases} \phantom{2x_1 + {}} x_2 - 2x_3 = -3 \\ 2x_1 + 3x_2 - x_3 = 1 \\ 6x_1 + 7x_2 + x_3 = 11 \end{cases} \tag{1.19}$$

Para resolver este sistema, seguindo a estratégia anterior, escalonamos a matriz completa associada:

$$\left[\begin{array}{ccc|c} 0 & 1 & -2 & -3 \\ \boxed{2} & 3 & -1 & 1 \\ 6 & 7 & 1 & 11 \end{array}\right] \xrightarrow{\ell_1 \leftrightarrow \ell_2} \left[\begin{array}{ccc|c} \boxed{2} & 3 & -1 & 1 \\ 0 & 1 & -2 & -3 \\ 6 & 7 & 1 & 11 \end{array}\right] \longrightarrow$$

$$\xrightarrow{\ell_3 \to \ell_3 - 3\ell_1} \left[\begin{array}{ccc|c} \boxed{2} & 3 & -1 & 1 \\ 0 & \boxed{1} & -2 & -3 \\ 0 & -2 & 4 & 8 \end{array}\right] \xrightarrow{\ell_3 \to \ell_3 + 2\ell_2} \left[\begin{array}{ccc|c} \boxed{2} & 3 & -1 & 5 \\ 0 & \boxed{1} & -2 & -3 \\ 0 & 0 & 0 & \boxed{2} \end{array}\right]$$

Obtivemos uma forma escalonada, e, neste caso, não é necessário chegar à forma escalonada reduzida. Observe que a terceira linha da matriz escalonada representa a equação $0 = 2$ ($0x_1 + 0x_2 + 0x_3 = 2$). Esta equação jamais poderá ser satisfeita, não interessa quais sejam os valores de x_1, x_2 e x_3. Portanto, o sistema (1.19) é *impossível*, isto é, não tem solução. Em outras palavras, o seu conjunto-solução é vazio.

Resolvemos, portanto, os sistemas (1.16) e (1.19). No exemplo 1.7, obtivemos uma descrição do conjunto-solução do primeiro, e, no exemplo 1.8, concluímos que o segundo sistema não tem solução alguma.

Vamos, a seguir, estudar mais cuidadosamente descrições tais como (1.18), formalizando os conceitos de variáveis "básicas" e "livres". Na subseção 1.7.2, abordaremos a questão fundamental de determinar se um dado sistema tem ou não solução. Por ora, chamamos a atenção do leitor para a equação "problemática" $0 = 2$, que encontramos no exemplo acima. Ela é típica de um sistema impossível. Aprofundaremos esta questão na seção 1.8.

1.7.1 Descrições paramétricas de conjuntos-solução, variáveis básicas e livres

Descrições do tipo (1.18) ou (1.10), na página 9, em que "variáveis básicas" são dadas em termos de "variáveis livres", são chamadas de **descrições paramétricas** de conjuntos-solução, pois as variáveis livres também podem ser chamadas de "parâmetros livres".

Reveja a passagem de (1.17) para (1.18) no exemplo 1.7. Poderíamos reformular (1.18) escolhendo, por exemplo, x_1 e x_3 como variáveis livres, e escrevendo x_2 e x_4 em termos delas.[5] Poderíamos, também, escolher x_1 e x_4 como variáveis

[5] Isso iria requerer mais algumas contas simples.

livres, e escrever x_2 e x_3 em termos delas. No entanto, x_3 e x_4 nunca poderiam ser *simultaneamente* livres, pois uma está amarrada à outra pela relação $x_3 - 3x_4 = 7$.

Há uma certa liberdade de escolha, portanto, ao classificarmos variáveis como dependentes ou livres. A escolha que adotamos, por *convenção*, é que as **variáveis básicas**, ou **dependentes**, sejam *aquelas que correspondam às colunas-pivô da matriz de coeficientes* de um dado sistema, e as demais sejam as **variáveis livres**. Lembre que é necessário escalonar uma matriz para que se possa "desvendar" quais são as colunas-pivô!

Vejamos, novamente, o exemplo 1.7. As colunas-pivô da matriz completa G são a primeira e a terceira, como fica claro ao examinar a matriz H, que é uma forma escalonada de G. Estabelecemos, portanto, que as variáveis x_1 e x_3 são básicas. As variáveis x_2 e x_4 correspondem à segunda e à quarta coluna de G, respectivamente, que são colunas não-pivô. Estas variáveis, portanto, são livres.

A quinta coluna de G é não-pivô, mas *não corresponde a uma variável livre*. De fato, a última coluna de G *não corresponde a variável alguma*: esta é a coluna associada aos termos independentes do sistema (1.16) (ver seção 1.2).

A classificação de variáveis em básicas ou livres só tem sentido no caso de sistemas possíveis. Portanto, não faz sentido classificar as variáveis do sistema (1.19). Isso porque o conjunto-solução de um sistema impossível é vazio, portanto, não tem descrição paramétrica.

Vamos examinar mais dois exemplos para fixar essas ideias.

Exemplo 1.9

Classifique cada variável do sistema abaixo como básica ou livre, e obtenha uma descrição paramétrica de seu conjunto-solução.

$$\begin{cases} - 3x_2 - 6x_3 + 4x_4 = 9 \\ -x_1 - 2x_2 - x_3 + 3x_4 = 1 \\ -2x_1 - 3x_2 + 3x_4 = -1 \\ x_1 + 4x_2 + 5x_3 - 9x_4 = -7 \end{cases} \tag{1.20}$$

Solução: Usando a mesma estratégia do exemplo 1.7, começamos escalonando a matriz completa associada ao sistema. Observe que a matriz completa de (1.20) é exatamente a matriz A de (1.13), na página 13. Já aplicamos o algoritmo de escalonamento a A na seção 1.5, e obtivemos a matriz B, dada em (1.15):

$$A = \begin{bmatrix} 0 & -3 & -6 & 4 & 9 \\ -1 & -2 & -1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ 1 & 4 & 5 & -9 & -7 \end{bmatrix} \xrightarrow[\text{(ver seção 1.5)}]{\text{escalonamento}} B = \begin{bmatrix} \boxed{1} & 0 & -3 & 0 & 5 \\ 0 & \boxed{1} & 2 & 0 & -3 \\ 0 & 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Assim, o sistema (1.20) é equivalente ao sistema associado a B:

$$\begin{cases} x_1 - 3x_3 = 5 \\ x_2 + 2x_3 = -3 \\ x_4 = 0 \end{cases} \tag{1.21}$$

Acima, descartamos a equação $0 = 0$ representada pela última linha de B.

As variáveis x_1, x_2 e x_4 correspondem às colunas-pivô da matriz A (ver seção 1.6, em particular o exemplo 1.6), portanto, estabelecemos que estas são as variáveis *básicas*. A variável x_3 corresponde a uma coluna não-pivô de A, portanto, x_3 é *livre*.

Escrevendo as variáveis básicas x_1, x_2 e x_4 em termos da variável livre x_3, obtemos uma descrição paramétrica do conjunto-solução de (1.20):

$$\begin{cases} x_1 = 5 + 3x_3, \\ x_2 = -3 - 2x_3, \\ x_3 \text{ é uma variável livre}, \\ x_4 = 0. \end{cases} \tag{1.22}$$

O sistema (1.20), mais uma vez, tem uma infinidade de soluções: uma para cada valor arbitrado ("chute") para variável livre x_3. □

Nem todo sistema linear, ainda que possível, tem variáveis livres.

Exemplo 1.10
Vamos revisitar, brevemente, o "primeiro exemplo" da seção 1.2 (página 4):

$$\begin{cases} x_1 - 4x_2 + x_3 = 0 \\ 3x_2 - 6x_3 = -3 \\ -3x_1 + 14x_2 - 6x_3 = 2 \end{cases} \tag{1.23}$$

Vimos que a matriz completa deste sistema e sua forma escalonada reduzida são dadas por

$$\left[\begin{array}{rrr|r} 1 & -4 & 1 & 0 \\ 0 & 3 & -6 & -3 \\ -3 & 14 & -6 & 2 \end{array}\right] \xrightarrow[\text{(ver seção 1.2)}]{\text{escalonamento}} \left[\begin{array}{rrr|r} \boxed{1} & 0 & 0 & 24 \\ 0 & \boxed{1} & 0 & 7 \\ 0 & 0 & \boxed{1} & 4 \end{array}\right]. \tag{1.24}$$

Todas as colunas da *matriz de coeficientes* do sistema são colunas-pivô, isto é, as colunas associadas a x_1, x_2 e x_3 são colunas-pivô (a última coluna da matriz completa *não* faz parte da matriz de coeficientes). Por esta razão, não há variáveis livres neste exemplo, e o sistema (1.23) possui *uma única solução*, dada por

$$\begin{cases} x_1 = 24, \\ x_2 = 7, \\ x_3 = 4. \end{cases} \tag{1.25}$$

Podemos chamar (1.25) de descrição paramétrica do conjunto-solução do sistema (1.23), apesar de a terminologia ser um tanto artificiosa neste caso. Como o sistema (1.23) não tem variáveis livres, não há "parâmetro" algum em (1.25).

1.7.2 Sistemas possíveis *versus* impossíveis

Nossa estratégia para resolver o sistema (1.16) do exemplo 1.7 foi a seguinte:

(a) escrever a matriz completa G associada ao sistema;
(b) obter sua forma escalonada reduzida H;
(c) extrair de H uma descrição paramétrica do conjunto-solução.

Usamos a mesma estratégia nos exemplos 1.9 e 1.10. Os sistemas destes três exemplos são todos *possíveis*. Todos têm ao menos uma solução. Note que o sistema (1.23) tem exatamente uma solução, e os sistemas (1.16) e (1.20) têm uma infinidade.

Por outro lado, ao escalonar a matriz completa do sistema (1.19) do exemplo 1.8, obtivemos uma linha igual a $\left[\begin{array}{ccc|c} 0 & 0 & 0 & 2 \end{array}\right]$, que representa a equação "problemática" $0x_1 + 0x_2 + 0x_3 = 2$ (ou $0 = 2$). Concluímos, imediatamente, que o sistema em questão é *impossível*: não tem solução alguma.

Isso motiva o seguinte critério para determinar se um dado sistema é ou não possível.

Proposição 1.11
Se uma forma escalonada da matriz completa de um sistema linear tem uma linha do tipo

$$\begin{bmatrix} 0 & 0 & \cdots & 0 & \blacksquare \end{bmatrix}, \quad \textit{com } \blacksquare \textit{ diferente de zero,} \tag{1.26}$$

então o sistema é impossível. Se, do contrário, uma forma escalonada da matriz completa não *tem nenhuma linha deste tipo, então o sistema é possível.*

Demonstração: Se uma forma escalonada da matriz completa de um sistema tem uma linha do tipo acima, podemos imediatamente concluir que tal sistema é impossível. Essa linha representa a equação $0 = \blacksquare$, que jamais será satisfeita (é uma equação do tipo $0 = 2$).

Por outro lado, se uma forma escalonada da matriz completa de um sistema *não* tem uma linha do tipo acima, então podemos, sem nenhum impedimento, determinar o valor de cada variável básica ou escrevê-la em termos das variáveis livres (se houver alguma). O resultado será uma descrição paramétrica do conjunto-solução. $\square$

O método recomendado para obter uma descrição paramétrica do conjunto-solução, no caso de um sistema possível, é aquele delineado no início desta subseção: passar à forma escalonada *reduzida* da matriz completa, como fizemos nos exemplos 1.7, 1.9 e 1.10. O procedimento geral é resumido na subseção abaixo.

1.7.3 Procedimento para a resolução de sistemas

Lembre que *resolver* um sistema linear significa obter uma descrição de seu conjunto-solução (uma descrição paramétrica, por exemplo), ou, então, concluir que o sistema é impossível. Abaixo, apresentamos um procedimento sistemático (um algoritmo) para resolver um dado sistema. Para estudá-lo, use os exemplos anteriores desta seção como guias.

Passo 1: Escreva a matriz completa do sistema.

Passo 2: Obtenha uma forma escalonada da matriz completa, usando o algoritmo visto na subseção 1.5.1. Não é necessário, por ora, chegar à forma reduzida.

Passo 3: Se houver uma linha do tipo (1.26) na matriz obtida, o sistema é impossível (conforme a proposição 1.11). Neste caso, o procedimento termina aqui. Caso contrário, siga para o próximo passo.

Passo 4: Obtenha a forma escalonada reduzida da matriz completa, usando o algoritmo visto na subseção 1.5.2.

Passo 5: Escreva o sistema de equações representado pela matriz obtida no passo anterior. (Este passo é dispensável. Com um pouco de prática, é fácil passar diretamente ao passo 6, pulando o 5.)

Passo 6: Obtenha uma descrição paramétrica do conjunto-solução. Isso é feito reescrevendo cada equação não-nula (que não seja do tipo $0 = 0$) do passo anterior, de forma que a variável básica envolvida seja expressa em termos das variáveis livres (se houver alguma).

1.8 Existência e unicidade de solução

Nesta seção, abordaremos o resultado anunciado na introdução do capítulo: um sistema linear qualquer ou tem exatamente uma solução, ou não tem solução alguma, ou tem uma infinidade de soluções. Isso já foi praticamente demonstrado na seção anterior. Só falta sistematizar as ideias e enunciar os resultados.

Observação
Queremos fazer um breve comentário sobre terminologia. Dada uma classe de problemas matemáticos (sistemas lineares, por exemplo), duas questões naturais, frequentemente, se colocam: Sob quais condições um problema particular da classe terá alguma solução? Nos casos em que há solução, sob quais condições adicionais (talvez, mais restritas) haverá uma única solução, e sob quais condições haverá mais de uma? A primeira questão é chamada de um *problema de existência*, já que queremos saber quando é que uma solução existe. A segunda chama-se um *problema de unicidade*, já que queremos saber quando é que uma solução tem a qualidade de ser única. O mundo matemático é cheio de "teoremas de existência e unicidade", referentes aos mais variados problemas. Ao final desta seção, veremos um exemplo referente ao caso de sistemas lineares.

1.8.1 Existência

Na proposição 1.11 da seção anterior, estabelecemos um critério para determinar se um sistema possui solução. Tal critério baseia-se na (não) ocorrência de uma "linha problemática" do tipo (1.26), em uma forma escalonada da matriz completa do sistema.

Este já é um critério satisfatório para determinar a existência, ou não, de uma solução para um sistema linear. Mas desejamos reformulá-lo em termos das colunas-pivô da matriz completa do sistema.

Considere as matrizes escalonadas "genéricas"

$$\left[\begin{array}{ccc|c} \blacksquare & * & * & * \\ 0 & \blacksquare & * & * \\ 0 & 0 & \blacksquare & * \end{array}\right], \quad \left[\begin{array}{cccc|c} \blacksquare & * & * & * & * \\ 0 & 0 & \blacksquare & * & * \\ 0 & 0 & 0 & \blacksquare & * \end{array}\right] \quad \text{e} \quad \left[\begin{array}{cccccc|c} 0 & \blacksquare & * & * & * & * & * \\ 0 & 0 & \blacksquare & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & \blacksquare & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right], \tag{1.27}$$

e interprete cada uma como a matriz completa de um sistema linear (ou linha-equivalente a tal matriz). As matrizes acima representam sistemas *possíveis*, já que nenhuma tem uma linha do tipo (1.26). Já as matrizes

$$\left[\begin{array}{cc|c} \blacksquare & * & * \\ 0 & \blacksquare & * \\ 0 & 0 & \blacksquare \end{array}\right], \quad \left[\begin{array}{cccc|c} \blacksquare & * & * & * & * \\ 0 & 0 & \blacksquare & * & * \\ 0 & 0 & 0 & 0 & \blacksquare \end{array}\right] \quad \text{e} \quad \left[\begin{array}{cccccc|c} 0 & \blacksquare & * & * & * & * & * \\ 0 & 0 & \blacksquare & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & \blacksquare & * \\ 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \tag{1.28}$$

correspondem a sistemas *impossíveis*, já que elas possuem linhas do tipo (1.26), representando equações do tipo $0 = \blacksquare$. Lembre que $\blacksquare$ representa um número diferente de zero. Em cada matriz, a "linha problemática" está hachurada.

Examine as matrizes acima, dando atenção especial à *última coluna* de cada uma. As matrizes de (1.27) *não* possuem "linhas problemáticas", e, em cada uma delas, a última coluna é *não*-pivô. Em contrapartida, todas as matrizes de (1.28) apresentam "linhas problemáticas", e, em cada uma, a última coluna *é* pivô. O exercício P1.13 pede que você deduza o seguinte fato geral: *A última coluna de uma matriz qualquer* (*escalonada ou não*) *é pivô se e somente se uma de suas formas escalonadas tem uma linha do tipo* (1.26).

A proposição abaixo é completamente equivalente à proposição 1.11, em vista dessas observações.

Proposição 1.12
Um sistema linear é possível se e somente se a última coluna de sua matriz completa (*a dos termos independentes*) não *é uma coluna-pivô.*

Enfatizamos que, para determinar se um sistema é possível, *não* é necessário obter a forma escalonada *reduzida* de sua matriz completa. Basta chegar a uma forma escalonada *qualquer* para aplicar o critério da proposição acima.

1.8.2 Unicidade

Vimos que o sistema do exemplo 1.9 tem uma infinidade de soluções: cada valor que for escolhido (ou "chutado") para a variável livre x_3 leva a uma solução diferente. O exemplo 1.7 é semelhante, mas, neste caso, há *duas* variáveis livres: obtém-se uma solução diferente do sistema (1.16) para cada escolha de valores das variáveis x_2 e x_4.

O sistema do exemplo 1.10, por outro lado, não possui variáveis livres, e, assim, o sistema tem uma única solução. Em outras palavras, o conjunto-solução tem apenas um elemento.

Tendo esses exemplos por base, não é difícil formular as seguintes ideias gerais. Um sistema linear possível *ou tem pelo menos uma variável livre, ou então não tem variável livre alguma*. Claramente, não há outra alternativa. Se o sistema *tiver* pelo menos uma variável livre, então ele terá uma infinidade de soluções: uma para cada "chute" da(s) variável(eis) livre(s). Se o sistema *não tiver* nenhuma variável livre, então ele terá uma única solução. Salientamos que esta discussão refere-se, por hipótese, a um sistema *possível*.

Considere, por exemplo, as matrizes escalonadas

$$\left[\begin{array}{ccc|c} \blacksquare & \ast & \ast & \ast \\ 0 & \blacksquare & \ast & \ast \\ 0 & 0 & \blacksquare & \ast \end{array}\right], \quad \left[\begin{array}{ccc|c} \blacksquare & \ast & \ast & \ast \\ 0 & \blacksquare & \ast & \ast \\ 0 & 0 & \blacksquare & \ast \\ 0 & 0 & 0 & 0 \end{array}\right], \quad \left[\begin{array}{cc|c} \blacksquare & \ast & \ast \\ 0 & \blacksquare & \ast \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right] \quad \text{e} \quad \left[\begin{array}{cccc|c} \blacksquare & \ast & \ast & \ast & \ast \\ 0 & \blacksquare & \ast & \ast & \ast \\ 0 & 0 & \blacksquare & \ast & \ast \\ 0 & 0 & 0 & \blacksquare & \ast \end{array}\right], \tag{1.29}$$

e, mais uma vez, interprete cada uma como a matriz *completa* de um sistema linear. As matrizes acima representam sistemas possíveis que *não* têm variáveis livres: Em cada caso, todas as colunas da matriz de coeficientes contêm uma posição-pivô, portanto, *todas as variáveis são básicas* (veja a subseção 1.7.1). Acima, a última coluna de cada matriz completa é não-pivô, mas *não* faz parte da matriz de *coeficientes*. Lembre-se de que a linha vertical destaca a matriz de coeficientes da coluna dos termos independentes.

Em contrapartida, as matrizes

$$\left[\begin{array}{ccc|c} \blacksquare & \ast & \ast & \ast \\ 0 & \blacksquare & \ast & \ast \\ 0 & 0 & 0 & 0 \end{array}\right], \quad \left[\begin{array}{ccc|c} \blacksquare & \ast & \ast & \ast \\ 0 & 0 & \blacksquare & \ast \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right], \quad \left[\begin{array}{cc|c} 0 & \blacksquare & \ast \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right] \quad \text{e} \quad \left[\begin{array}{cccc|c} 0 & \blacksquare & \ast & \ast & \ast \\ 0 & 0 & \blacksquare & \ast & \ast \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \tag{1.30}$$

representam sistemas possíveis *que têm variáveis livres*. As colunas associadas às variáveis livres estão hachuradas. Elas são as colunas não-pivô em cada matriz de *coeficientes* (ver subseção 1.7.1). No sistema representado pela primeira matriz acima, por exemplo, a variável x_3 é livre. No caso do sistema representado pela última matriz, as variáveis livres são x_1 e x_4. Note que, em cada uma dessas matrizes, a última coluna não está hachurada. Apesar de serem não-pivô, essas colunas *não correspondem a variáveis livres*! Elas não correspondem a variável alguma, já que, repetimos, não fazem parte da matriz de coeficientes.

Para determinar se um sistema possível tem variáveis livres, portanto, basta determinar se há colunas não-pivô em sua matriz de coeficientes.

Por fim, as matrizes

$$\left[\begin{array}{cc|c} \blacksquare & \ast & \ast \\ 0 & \blacksquare & \ast \\ 0 & 0 & \blacksquare \end{array}\right], \quad \left[\begin{array}{cccc|c} \blacksquare & \ast & \ast & \ast & \ast \\ 0 & 0 & \blacksquare & \ast & \ast \\ 0 & 0 & 0 & 0 & \blacksquare \end{array}\right] \quad \text{e} \quad \left[\begin{array}{cccc|c} \blacksquare & \ast & \ast & \ast & \ast \\ 0 & \blacksquare & \ast & \ast & \ast \\ 0 & 0 & 0 & 0 & \blacksquare \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \tag{1.31}$$

representam sistemas *impossíveis*: as linhas hachuradas são "problemáticas". Observe, também, que, em cada matriz, a última coluna é pivô. A classificação das variáveis em básicas ou livres é irrelevante nestes casos.

Em resumo:

- As matrizes em (1.29) representam sistemas com solução única (sistemas possíveis, sem variáveis livres). Em outras palavras, o conjunto-solução de cada um desses sistemas tem um único elemento.
- As matrizes em (1.30) representam sistemas com uma infinidade de soluções (sistemas possíveis, com variáveis livres). O conjunto-solução de cada um desses sistemas tem uma infinidade de elementos.
- As matrizes em (1.31) representam sistemas sem solução alguma (sistemas impossíveis). O conjunto-solução de cada um deles é vazio, ou seja, não tem elemento algum.

1.8.3 Teorema de existência e unicidade de solução

Sintetizamos, abaixo, os resultados das duas últimas subseções. Se você achar confuso o enunciado do teorema abaixo, refira-se ao exercício P1.14.

Teorema 1.13
Um sistema linear é possível se e somente se a última coluna de sua matriz completa não *é uma coluna-pivô* (*ver proposição 1.12*).

Um sistema possível tem uma única solução se e somente se todas as colunas de sua matriz de coeficientes *são colunas-pivô* (*neste caso, não há variáveis livres: todas as variáveis do sistema são básicas*).

Se, do contrário, a matriz de coeficientes de um sistema possível tem, pelo menos, uma coluna não-pivô, então o sistema tem, pelo menos, uma variável livre, e, portanto, uma infinidade de soluções.

Um sistema linear, portanto, ou tem uma única solução, ou uma infinidade, ou nenhuma. Para determinar em qual caso enquadra-se um dado sistema, basta escalonar sua matriz completa e usar o teorema acima. O escalonamento é necessário para desvendar quais são as colunas-pivô da matriz completa. Observe, mais uma vez, que basta obter uma forma escalonada *qualquer*: não é necessário chegar à forma reduzida. Ou seja, não é necessário aplicar todos os passos do procedimento da subseção 1.7.3, basta ir até o passo 2.

Observação
Talvez, você ainda não tenha percebido todas as consequências do teorema acima, e sua importância na análise de sistemas lineares. O teorema implica, por exemplo, que não há sistemas lineares com *exatamente* duas, três ou, digamos, dezessete soluções (um sistema linear ou tem uma solução, ou uma infinidade, ou nenhuma).

A situação é muito diferente no caso de sistemas não-lineares. A equação não-linear $x^2 = 4$, por exemplo, tem exatamente duas soluções: $x = 2$ e $x = -2$. E a equação $(x-1)(x-2)(x-3)\cdots(x-17) = 0$ tem dezessete soluções (quais são?).

Exercícios propostos

P1.1. Verifique que o sistema (b) da introdução não possui soluções, e que o sistema (c) possui uma infinidade delas.

P1.2. Verifique que as equações do sistema (c) da introdução representam a mesma reta no plano (x_1, x_2).

P1.3. A primeira equação do sistema (a) da introdução corresponde a qual das duas retas da figura 1.1(a)?

P1.4. Determine o ponto de interseção entre as retas $2x_1+x_2 = 1$ e $x_1-2x_2 = 13$.

P1.5. Descreva a interseção entre as retas $3x_1 - 2x_2 = 4$ e $6x_1 - 4x_2 = 8$. A interseção é vazia? É um ponto? É uma reta?

P1.6. Determine quais das matrizes abaixo estão em forma escalonada. Quais estão na forma escalonada reduzida?

(a) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

(d) $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ (e) $\begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ (f) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$

(g) $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ (h) $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ (i) $\begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 \end{bmatrix}$

(j) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ (k) $\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ (l) $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 5 & 6 & 7 \\ 0 & 0 & 0 & 8 \end{bmatrix}$

P1.7. Encontre uma forma escalonada de cada matriz abaixo. Indique as posições e as colunas-pivô. *Observação*: Não é necessário obter a forma reduzida.

(a) $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 9 & 12 & 15 \end{bmatrix}$

(e) $\begin{bmatrix} 2 & 1 & 4 & 4 & 17 \\ 2 & -3 & 12 & -1 & -2 \\ -2 & -3 & 0 & 1 & -4 \\ 1 & 0 & 3 & -2 & -4 \end{bmatrix}$ (f) $\begin{bmatrix} 1 & -2 & -3 & -7 & -19 \\ 0 & 0 & 2 & 4 & 16 \\ -4 & 8 & -2 & 0 & -36 \\ -1 & 2 & 0 & 1 & -5 \end{bmatrix}$

P1.8. Obtenha, agora, a forma escalonada reduzida de cada matriz do exercício anterior. Aproveite o trabalho já realizado.

P1.9. Em cada matriz abaixo, a posição destacada é uma posição-pivô? Justifique cada resposta.

(a) $\begin{bmatrix} -3 & -9 & 3 & -9 \\ 4 & 12 & \boxed{0} & 20 \\ 1 & 3 & 5 & 15 \end{bmatrix}$ (b) $\begin{bmatrix} 3 & 5 & -11 \\ -3 & \boxed{-5} & 17 \\ -6 & -10 & -2 \end{bmatrix}$

P1.10. Determine quais dos sistemas abaixo são possíveis. Para cada sistema possível, classifique as variáveis em básicas ou livres, e forneça uma descrição paramétrica do conjunto-solução. *Use o procedimento da subseção 1.7.3.*

(a) $\begin{cases} \phantom{x_1 + {}} x_2 - 2x_3 = 7 \\ x_1 + x_2 + 3x_3 = 11 \\ x_1 + 3x_2 - x_3 = 25 \end{cases}$

(b) $\begin{cases} 4x_1 - 3x_2 + 26x_3 = -5 \\ -x_1 + x_2 - 7x_3 = 3 \\ \phantom{-x_1 + {}} 2x_2 - 4x_3 = 10 \end{cases}$

(c) $\begin{cases} -4x_1 + 4x_2 + 3x_3 - 22x_4 = -27 \\ -5x_1 + 5x_2 - 4x_3 - 12x_4 = 5 \\ 2x_1 - 2x_2 \phantom{{}- 4x_3} + 8x_4 = 6 \end{cases}$

P1.11. Pense em cada matriz do exercício P1.7 como a matriz completa de um sistema. Escreva cada um desses sistemas, e resolva-o. *Dica*: Aproveite o trabalho já realizado no exercício P1.8.

P1.12. Considere o seguinte sistema de equações lineares:

$$\begin{cases} x_1 + hx_2 = 2 \\ 3x_1 - 2x_2 = k \end{cases}$$

Determine os valores de h e k tais que o sistema tenha (i) nenhuma solução, (ii) uma única solução, e (iii) muitas soluções. Dê respostas separadas para cada parte, justificando cada uma.

P1.13. Mostre que a última coluna de uma matriz qualquer é uma coluna-pivô se e somente se uma de suas formas escalonadas tem uma linha do tipo (1.26).

P1.14. Justifique as seguintes afirmativas, usando o teorema 1.13. (Este exercício é, simplesmente, uma reformulação desse teorema.)

(a) Um sistema linear *não terá solução alguma* se e só se houver uma posição-pivô na última coluna de sua matriz completa.

(b) Um sistema linear terá *uma única solução* se e só se houver uma posição-pivô em todas as colunas de sua matriz completa, *exceto na última coluna*.

(c) Um sistema linear terá *uma infinidade de soluções* se não houver posição-pivô na última coluna de sua matriz completa e, além disso, não houver posição-pivô em alguma outra coluna.

P1.15. Determine se cada afirmativa é verdadeira ou falsa. Justifique.

(a) Nem toda matriz possui uma forma escalonada.
(b) Nem toda matriz possui uma forma escalonada reduzida.
(c) Uma matriz pode ter muitas formas escalonadas reduzidas distintas.
(d) Uma matriz pode ter muitas formas escalonadas distintas.
(e) Uma matriz pode ter uma única forma escalonada. *Dica*: Veja o exercício P1.6(k).

P1.16. Mostre que, se existe uma sequência de operações-linha que leva a matriz A à matriz B, então existe uma sequência que leva B a A. *Dica*: Use a proposição 1.2.

P1.17. Se uma matriz A é linha-equivalente a B, e B é linha-equivalente a C, então A e C são linha-equivalentes. Justifique essa afirmativa.

P1.18. As matrizes dos itens (c) e (d) do exercício P1.7 são linha-equivalentes? Justifique. *Dica*: Use os dois exercícios anteriores, e aproveite o trabalho já realizado no exercício P1.8.

P1.19. Justifique a seguinte afirmativa: O número de posições-pivô de uma matriz não pode exceder o seu número de linhas, nem o de colunas.

Dicas: Reveja a definição de *posição-pivô* na seção 1.6, e também a definição 1.5, na página 10. Agora, reflita: Uma linha de uma matriz qualquer pode ter mais do que um elemento líder? Uma matriz escalonada pode ter dois elementos líderes na mesma coluna?

P1.20. Convença-se de que as únicas quatro formas escalonadas "genéricas" de tamanho 2×2 são

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} \blacksquare & * \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & \blacksquare \\ 0 & 0 \end{bmatrix} \quad \text{e} \quad \begin{bmatrix} \blacksquare & * \\ 0 & \blacksquare \end{bmatrix}.$$

P1.21. Existem exatamente 11 formas escalonadas "genéricas" de tamanho 2×4. Escreva cada uma delas. *Dica*: Escreva aquela(s) com *nenhuma* posição-pivô, em seguida aquela(s) com *uma*, e finalmente aquela(s) com *duas*. Por que é que o processo termina por aí?

P1.22. Mostre que um sistema linear com duas equações e três variáveis ou não tem solução alguma, ou tem uma infinidade de soluções. *Dica*: Use o exercício anterior.

Capítulo 2

Vetores e Combinações Lineares

2.0 Introdução

Neste capítulo, vamos estabelecer as "pedras fundamentais" da álgebra linear. Definiremos vetores, as operações de soma de vetores e produto por escalar e o conceito importantíssimo de *combinação linear*. Em certo sentido, podemos dizer que a teoria de álgebra linear começa aqui. Usaremos a teoria do capítulo anterior ocasionalmente, mas, neste capítulo, sistemas lineares deixarão de ter papel central.

A última seção deste capítulo contém uma discussão interessante, que será generalizada no capítulo 4. Dado um "espaço ambiente", caracterizamos os subconjuntos de vetores que são capazes de "gerar", por combinações lineares, todos os outros vetores do espaço. No decorrer do capítulo, o significado disso ficará mais claro, e a importância desses "conjuntos geradores" ficará evidente ao longo do curso.

2.1 Vetores de $\mathbb{R}^n$

Um **vetor de** n **coordenadas** é uma lista ordenada de n números, que, usualmente, escrevemos em uma coluna. As listas

$$\mathbf{u} = \begin{bmatrix} 5 \\ -1 \end{bmatrix} \qquad \text{e} \qquad \mathbf{v} = \begin{bmatrix} -3 \\ 1/5 \\ \sqrt{17} \end{bmatrix},$$

por exemplo, são vetores de duas e de três coordenadas, respectivamente. Pode-se pensar em um vetor como uma matriz de apenas uma coluna, com a ressalva de que é mais usual chamar os números em um vetor de *componentes* ou *coordenadas* (ao invés de *entradas* ou *elementos*, que é como chamamos os números em uma matriz, como vimos no capítulo 1).

Este texto tratará quase exclusivamente de vetores cujas coordenadas são *números reais*, mas todos os conceitos e resultados se generalizam para vetores de coordenadas complexas.[1] Vetores e matrizes de números complexos têm grande importância em algumas áreas da física e da engenharia.

2.1.1 Notação

Números reais (ou complexos) em álgebra linear são, usualmente, chamados de **escalares**. O conjunto dos números reais é denotado por $\mathbb{R}$, e o dos complexos por $\mathbb{C}$. Quando dizemos que "c é um escalar", nossa intenção, geralmente, é enfatizar que c é um *número*, e não um vetor.

O conjunto de todos os vetores de n componentes reais é denotado por $\mathbb{R}^n$. O "$\mathbb{R}$" indica que as componentes são números reais, e o "n", o número de componentes. Em contrapartida, o conjunto dos vetores de n componentes *complexas* é denotado por $\mathbb{C}^n$.

O vetor $\mathbf{u}$ dado acima, por exemplo, é um elemento de $\mathbb{R}^2$, enquanto $\mathbf{v}$ é um elemento de $\mathbb{R}^3$. Em símbolos, escrevemos $\mathbf{u} \in \mathbb{R}^2$ e $\mathbf{v} \in \mathbb{R}^3$ (lê-se "$\mathbf{u}$ pertence a erre-dois" e "$\mathbf{v}$ pertence a erre-três").

Neste texto, vetores serão denotados usando letras em negrito, como acima com $\mathbf{u}$ e $\mathbf{v}$. Escalares geralmente serão representados por letras *minúsculas* em tipo itálico ou letras gregas. Matrizes serão representadas por letras *maiúsculas* e, também, em tipo itálico. Escreveremos, por exemplo,

$$c = 3, \qquad \alpha = -5, \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \end{bmatrix} \quad \text{e} \quad A = \begin{bmatrix} 1 & 2 \\ -3 & 0 \end{bmatrix}.$$

As componentes x_1 e x_2 do vetor $\mathbf{x}$ são indicadas em itálico, pois são escalares, bem como os números c e α.

Chamamos o vetor de $\mathbb{R}^n$ cujas componentes são todas iguais a zero de **vetor zero**, ou de **vetor nulo**, e denotamo-lo por $\mathbf{0}_n$ (com o algarismo 0 em negrito). Assim, $\mathbf{0}_2 = \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]$ e $\mathbf{0}_3 = \left[\begin{smallmatrix} 0 \\ 0 \\ 0 \end{smallmatrix}\right]$, por exemplo. Quando não houver ambiguidade, podemos denotar o vetor zero de $\mathbb{R}^n$ por $\mathbf{0}$, simplesmente.[2]

A matriz de tamanho $n \times m$ cujas entradas são todas iguais a zero, chamada de **matriz zero** ou de **matriz nula**, será denotada por $\mathit{0}_{n\times m}$ (com o algarismo 0 em itálico), ou, simplesmente, por $\mathit{0}$, quando não houver ambiguidade quanto ao seu tamanho.

2.1.2 Dimensão de $\mathbb{R}^n$

A noção de "dimensão" é bastante usada fora do contexto matemático. Por exemplo, diversos filmes hollywoodianos, atualmente, são filmados em "3D" ("três dimensões"), e as diferenças entre exibições 2D e 3D são evidentes. Outro exemplo são as "imagens tridimensionais" do interior do corpo humano, obtidas via tomografia computadorizada ou ressonância magnética, que permitiram grandes

[1] Mas cuidado, porque, em alguns pontos da teoria, a generalização ao caso complexo requer atenção a certos detalhes. Alertaremos os leitores quando chegarmos a esses pontos.

[2] Mas cuidado para não fazer confusão, pois o vetor $\mathbf{0}$ de $\mathbb{R}^n$ *não* é a mesma coisa que o vetor $\mathbf{0}$ de $\mathbb{R}^m$, se $n \neq m$! O primeiro é uma lista de n zeros; o segundo, de m zeros.

avanços em medicina diagnóstica.[3] Em filmes de ficção científica que envolvem viagens no tempo, já é clichê ouvir algum personagem (geralmente um estereótipo de cientista) dizer que vivemos um em universo de "quatro dimensões": três dimensões espaciais e uma dimensão temporal.

Gostaríamos de definir precisamente o conceito de *dimensão*, no contexto matemático. Isso, porém, só poderá ser feito na seção 4.3, pois ainda não dispomos da teoria necessária. Por enquanto, *convencionamos dizer que a dimensão de* $\mathbb{R}^n$ *é o número inteiro não-negativo* n. Assim, por exemplo, a dimensão de $\mathbb{R}^2$ é 2, a dimensão de $\mathbb{R}^5$ é 5, e a dimensão de $\mathbb{R}^{1017}$ é 1017.

Veremos que esta convenção será coerente com a definição de dimensão, mais geral e rigorosa, da seção 4.3. Introduzimos a terminologia antecipadamente para podermos usá-la e para nos habituarmos a ela.

A convenção acima também condiz com nossa percepção intuitiva, coloquial, de dimensão. Como recordaremos na próxima subseção, o conjunto $\mathbb{R}^2$ pode ser representado geometricamente por um plano, que é um objeto que consideramos "bidimensional". Já o $\mathbb{R}^3$ pode ser representado pelo espaço "tridimensional".

Cuidado!
Enfatizamos que o conceito de dimensão é mais geral do que o apresentado acima, e será visto somente na seção 4.3. Não pense que, por ler meramente esta pequena subseção, você já domina o conceito! (Esse aviso pode ser útil para quem estiver estudando de última hora para uma prova...)

2.1.3 Representação geométrica

O leitor já deve estar familiarizado com o sistema de coordenadas cartesianas e com a representação de vetores de $\mathbb{R}^2$ ou de $\mathbb{R}^3$ como "setas" no plano ou no espaço, respectivamente. Sendo assim, não desenvolveremos essas ideias em detalhe neste texto. Teceremos apenas alguns comentários e apresentaremos exemplos, a título de recapitulação.

Na figura 2.1(a), o conjunto $\mathbb{R}^2$ é representado geometricamente por um *plano coordenado*. O sistema de coordenadas cartesianas nesse plano é determinado pelos eixos x_1 e x_2 indicados. Cada vetor $\left[\begin{smallmatrix} v_1 \\ v_2 \end{smallmatrix}\right]$ de $\mathbb{R}^2$ corresponde a um ponto de coordenadas v_1 e v_2 com relação aos eixos horizontal e vertical, respectivamente. É mais intuitivo e usual, no entanto, representar vetores como *setas* no plano, ao invés de pontos. Representamos, na figura 2.1(a), por exemplo, os vetores $\mathbf{u} = \left[\begin{smallmatrix} 5 \\ 2 \end{smallmatrix}\right]$, $\mathbf{v} = \left[\begin{smallmatrix} -2 \\ 5 \end{smallmatrix}\right]$ e $\mathbf{w} = \left[\begin{smallmatrix} 4 \\ -1 \end{smallmatrix}\right]$. Observe que os vetores são sempre representados por setas que partem da *origem* do sistema de coordenadas, ou seja, do ponto cujas coordenadas são iguais a zero. O vetor zero $\mathbf{0}$ é representado nessa figura simplesmente como um ponto na origem (já que não é possível desenhar uma seta retilínea que começa e termina no mesmo ponto...).

A representação geométrica de $\mathbb{R}^3$ é feita de forma análoga, mas, nesse caso, usa-se um sistema de coordenadas cartesianas em um espaço *tridimensional*. Na

[3] A construção de "imagens 3D" a partir de "fatias 2D" nesses exames, a propósito, envolve métodos matemáticos sofisticados. E, na base desses métodos, há bastante álgebra linear.

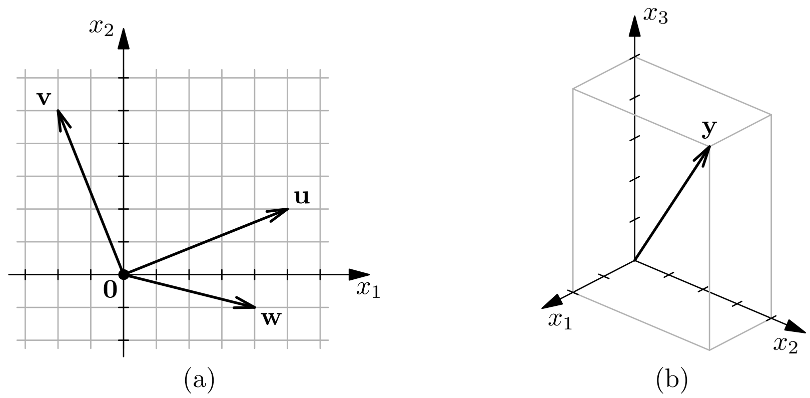


Figura 2.1: Representação geométrica de vetores em $\mathbb{R}^2$ (a) e em $\mathbb{R}^3$ (b).

figura 2.1(b), esse sistema é dado pelos três eixos indicados, e representamos geometricamente o vetor $\mathbf{y} = \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}$.

Em resumo, um vetor pode ser pensado como uma "seta partindo da origem" dentro de um "espaço ambiente" (um plano, no caso de $\mathbb{R}^2$; e um espaço tridimensional, no caso de $\mathbb{R}^3$). A representação é tão natural que, às vezes, dizemos coisas como "esta seta *é* o vetor $\mathbf{x}$", ao invés de "esta seta *representa* o vetor $\mathbf{x}$".

Observação

Surpreendentemente, a ideia da representação geométrica de vetores de $\mathbb{R}^n$ é útil mesmo quando estamos lidando com espaços de dimensão mais alta, isto é, quando n é maior do que três. Um vetor de $\mathbb{R}^7$, por exemplo, pode ser pensado como uma seta partindo da origem dentro de um "espaço ambiente de dimensão sete". Ninguém é capaz de visualizar diretamente um espaço de dimensão sete, então, isso pode parecer muito abstrato ou "viajante". No entanto, a imagem de vetores como "setas" pode ser elucidativa mesmo nesses casos. Logo adiante, na subseção 2.1.5, veremos um exemplo disso. Podemos conceber que os vetores representados nas figuras 2.2 e 2.3 pertençam a um espaço maior, e não apenas ao plano da página.

2.1.4 Igualdade entre vetores

Dizemos que dois vetores são **iguais** quando suas componentes correspondentes são todas iguais. Assim, se $\mathbf{u}$ e $\mathbf{v}$ são dois vetores de $\mathbb{R}^n$, então

$$\mathbf{u} = \mathbf{v} \quad \text{é uma forma sucinta de escrever} \quad \begin{cases} u_1 = v_1 \\ u_2 = v_2 \\ \quad\vdots \\ u_n = v_n, \end{cases} \tag{2.1}$$

onde $u_1, u_2, \ldots, u_n$ e $v_1, v_2, \ldots, v_n$ são as componentes dos vetores $\mathbf{u}$ e $\mathbf{v}$, respectivamente. Os vetores $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ e $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$ de $\mathbb{R}^2$ *não* são iguais (como dissemos, um vetor é uma lista *ordenada*).

Para que dois vetores $\mathbf{u}$ e $\mathbf{v}$ sejam *diferentes*, basta que *uma* das componentes de $\mathbf{u}$ seja diferente da componente correspondente de $\mathbf{v}$. Em outras palavras,

$\mathbf{u} \neq \mathbf{v}$ é uma forma resumida de dizer que *pelo menos uma* das igualdades à direita em (2.1) *não* vale. Mas isso não significa dizer que *todas* as componentes de $\mathbf{u}$ e $\mathbf{v}$ sejam diferentes.

Observe que a igualdade à esquerda, em (2.1), é uma *igualdade vetorial* (entre vetores), ao passo que as igualdades à direita são igualdades entre escalares (a que já estamos habituados).

2.1.5 Soma de vetores e produto de vetor por escalar

Dados dois vetores $\mathbf{u}$ e $\mathbf{v}$ de $\mathbb{R}^n$, definimos a **soma** $\mathbf{u}+\mathbf{v}$ como o vetor obtido por meio da soma das componentes correspondentes de $\mathbf{u}$ e $\mathbf{v}$, isto é,

$$\mathbf{u}+\mathbf{v} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{bmatrix} \in \mathbb{R}^n,$$

onde $u_1, u_2, \ldots, u_n$ e $v_1, v_2, \ldots, v_n$ são, mais uma vez, as componentes dos vetores $\mathbf{u}$ e $\mathbf{v}$, respectivamente. Enfatizamos que a soma $\mathbf{u}+\mathbf{v}$ é um elemento de $\mathbb{R}^n$, ou seja, do *mesmo* conjunto onde residem $\mathbf{u}$ e $\mathbf{v}$. Não faz sentido somar um vetor de $\mathbb{R}^n$ com um vetor de $\mathbb{R}^m$ quando $n \neq m$ (esta operação não é definida).

Por exemplo, se

$$\mathbf{u} = \begin{bmatrix} 2 \\ -3 \\ 0 \end{bmatrix} \quad \text{e} \quad \mathbf{v} = \begin{bmatrix} 1 \\ 3 \\ -5 \end{bmatrix}, \quad \text{então} \quad \mathbf{u}+\mathbf{v} = \begin{bmatrix} 2+1 \\ (-3)+3 \\ 0+(-5) \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ -5 \end{bmatrix}.$$

Todos os vetores acima são elementos de $\mathbb{R}^3$.

A soma de dois vetores é representada, geometricamente, pela "regra do paralelogramo", ilustrada na figura 2.2. Se $\mathbf{u}$ e $\mathbf{v}$ são representados pelas setas indicadas, que formam dois lados de um paralelogramo, então o vetor-soma $\mathbf{u}+\mathbf{v}$ é representado pela seta que corre ao longo da diagonal do paralelogramo. Lembre-se de que todos os vetores são representados por setas que partem da origem $\mathbf{0}$.

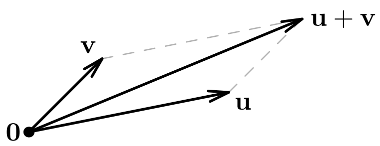


Figura 2.2: A "regra do paralelogramo" para a soma de dois vetores.

Dados um escalar $\alpha \in \mathbb{R}$ e um vetor $\mathbf{v} \in \mathbb{R}^n$, definimos o **produto** de $\mathbf{v}$ por α, denotado por $\alpha\mathbf{v}$, como o vetor obtido multiplicando cada componente de $\mathbf{v}$ por α, ou seja,

$$\alpha\mathbf{v} = \begin{bmatrix} \alpha v_1 \\ \alpha v_2 \\ \vdots \\ \alpha v_n \end{bmatrix} \in \mathbb{R}^n.$$

Assim, se

$$\alpha = -4 \quad \text{e} \quad \mathbf{v} = \begin{bmatrix} 1 \\ 3 \\ -5 \end{bmatrix}, \quad \text{então} \quad \alpha\mathbf{v} = (-4)\begin{bmatrix} 1 \\ 3 \\ -5 \end{bmatrix} = \begin{bmatrix} -4 \\ -12 \\ 20 \end{bmatrix}.$$

Geometricamente, o produto por α produz um "esticamento" de um vetor (se $|\alpha| > 1$) ou uma "contração" (se $|\alpha| < 1$). Se α é negativo, o produto $\alpha\mathbf{v}$ produz, ainda, uma "inversão no sentido" do vetor $\mathbf{v}$. A figura 2.3 ilustra um vetor $\mathbf{v}$ e seus múltiplos $3\mathbf{v}$, $\frac{1}{2}\mathbf{v}$ e $(-1)\mathbf{v}$. Podemos escrever $-\mathbf{v}$ ao invés de $(-1)\mathbf{v}$.

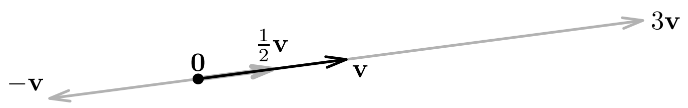


Figura 2.3: Representação geométrica do produto de um vetor por escalares.

As operações de soma de vetores e multiplicação por escalar podem ser combinadas de várias formas em uma expressão. Por exemplo, se $\mathbf{x} = \left[\begin{smallmatrix} 1 \\ -2 \end{smallmatrix}\right]$, $\mathbf{y} = \left[\begin{smallmatrix} 2 \\ 0 \end{smallmatrix}\right]$ e $\mathbf{z} = \left[\begin{smallmatrix} 3 \\ 1 \end{smallmatrix}\right]$, então

$$\begin{aligned} &2\big(3\mathbf{x} + (-1)\mathbf{y}\big) + 4\mathbf{z} = 2\left(3\begin{bmatrix} 1 \\ -2 \end{bmatrix} + (-1)\begin{bmatrix} 2 \\ 0 \end{bmatrix}\right) + 4\begin{bmatrix} 3 \\ 1 \end{bmatrix} = \\ &= 2\left(\begin{bmatrix} 3 \\ -6 \end{bmatrix} + \begin{bmatrix} -2 \\ 0 \end{bmatrix}\right) + \begin{bmatrix} 12 \\ 4 \end{bmatrix} = 2\begin{bmatrix} 1 \\ -6 \end{bmatrix} + \begin{bmatrix} 12 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ -12 \end{bmatrix} + \begin{bmatrix} 12 \\ 4 \end{bmatrix} = \begin{bmatrix} 14 \\ -8 \end{bmatrix}. \end{aligned} \tag{2.2}$$

Dados dois vetores $\mathbf{u}$ e $\mathbf{v}$, o vetor $\mathbf{u} + (-1)\mathbf{v}$ é obtido, como indicado, multiplicando $\mathbf{v}$ pelo escalar -1 e somando o resultado a $\mathbf{u}$. Chamamos esse vetor de *diferença* entre os vetores $\mathbf{u}$ e $\mathbf{v}$, e escrevemos $\mathbf{u} - \mathbf{v}$ ao invés de $\mathbf{u} + (-1)\mathbf{v}$ para simplificar a notação.

As operações que definimos acima têm boas propriedades algébricas. Traduzindo do "matematiquês": é fácil fazer e simplificar contas envolvendo as operaçoes que definimos acima, pois valem diversas propriedades "naturais" (ou "intuitivas").

Proposição 2.1 (Propriedades algébricas das operações vetoriais)
Para quaisquer vetores $\mathbf{u}$, $\mathbf{v}$ *e* $\mathbf{w}$ *de* $\mathbb{R}^n$ *e quaisquer escalares* α *e* β, *valem as propriedades abaixo.*

(a) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
(b) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
(c) $\mathbf{u} + \mathbf{0}_n = \mathbf{u}$
(d) $\mathbf{u} - \mathbf{u} = \mathbf{u} + (-1)\mathbf{u} = \mathbf{0}$
(e) $\alpha(\mathbf{u} + \mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v}$
(f) $(\alpha + \beta)\mathbf{u} = \alpha\mathbf{u} + \beta\mathbf{u}$
(g) $\alpha(\beta\mathbf{u}) = (\alpha\beta)\mathbf{u}$
(h) $1\mathbf{u} = \mathbf{u}$
(i) $0\mathbf{u} = \mathbf{0}_n$
(j) $\alpha\mathbf{0}_n = \mathbf{0}_n$

É fácil verificar as propriedades acima, e não perderemos muito tempo provando cada uma. Vejamos apenas a propriedade distributiva (e):

$$
\begin{aligned}
\alpha(\mathbf{u}+\mathbf{v}) &= \alpha \begin{bmatrix} u_1+v_1 \\ u_2+v_2 \\ \vdots \\ u_n+v_n \end{bmatrix} && \text{pela definição de soma de vetores,} \\
&= \begin{bmatrix} \alpha(u_1+v_1) \\ \alpha(u_2+v_2) \\ \vdots \\ \alpha(u_n+v_n) \end{bmatrix} && \text{pela definição de produto por escalar,} \\
&= \begin{bmatrix} \alpha u_1+\alpha v_1 \\ \alpha u_2+\alpha v_2 \\ \vdots \\ \alpha u_n+\alpha v_n \end{bmatrix} && \text{pela propriedade distributiva dos números reais (aplicada em cada coordenada),} \\
&= \begin{bmatrix} \alpha u_1 \\ \alpha u_2 \\ \vdots \\ \alpha u_n \end{bmatrix} + \begin{bmatrix} \alpha v_1 \\ \alpha v_2 \\ \vdots \\ \alpha v_n \end{bmatrix} && \text{pela definição de soma de vetores,} \\
&= \alpha\mathbf{u}+\alpha\mathbf{v} && \text{pela definição de produto por escalar.}
\end{aligned}
$$

Você pode verificar as outras propriedades por conta própria. A estratégia é sempre a mesma: escrever os vetores coordenada por coordenada, e usar as propriedades que você já conhece para números reais.

Em vista da propriedade (b), podemos escrever a soma $(\mathbf{u}+\mathbf{v})+\mathbf{w}$ simplesmente como $\mathbf{u}+\mathbf{v}+\mathbf{w}$, uma vez que a ordem em que somamos não importa. Do mesmo modo, uma soma da forma $\big((\mathbf{u}+\mathbf{v})+\mathbf{w}\big)+\mathbf{x}$ pode ser reescrita como $\mathbf{u}+\mathbf{v}+\mathbf{w}+\mathbf{x}$. Observações como essa simplificam muito os cálculos e a notação. A expressão $2\big(3\mathbf{x}+(-1)\mathbf{y}\big)+4\mathbf{z}$ em (2.2), por exemplo, pode ser reescrita como $6\mathbf{x}-2\mathbf{y}+4\mathbf{z}$ (usamos as propriedades (b), (e) e (g) para fazer esta simplificação).

As operações de soma de vetores e produto por escalar são as "pedras fundamentais" da álgebra linear. Todos os demais conceitos serão definidos, direta ou indiretamente, em termos dessas operações. Um exemplo é o conceito de *combinação linear*, que será o tema da próxima seção.

Observação
Por dispor das operações de soma, de produto por escalar, e das propriedades algébricas listadas acima, o conjunto $\mathbb{R}^n$ é um exemplo de *espaço vetorial*. Não definiremos esse conceito em toda sua generalidade, pois o nível de abstração exigido excederia os limites propostos por este texto. Recomendamos o livro de Paul Halmos [1] aos leitores interessados em um tratamento mais geral (e bem mais avançado!).

2.2 Combinações lineares

Nesta seção, faremos uma breve introdução a um dos conceitos cruciais de álgebra linear. Esse assunto será aprofundado na seção 2.5.

Definição 2.2
Dados os vetores $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ de $\mathbb{R}^n$ e os escalares $c_1, c_2, \ldots, c_m$, dizemos que o vetor

$$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_m\mathbf{a}_m \in \mathbb{R}^n \tag{2.3}$$

é a **combinação linear** de $\mathbf{a}_1, \ldots, \mathbf{a}_m$ com **pesos** (ou coeficientes) $c_1, \ldots, c_m$.

Por exemplo, se $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{a}_3$ são vetores de $\mathbb{R}^7$ (não interessa quais sejam suas coordenadas), então

$$6\mathbf{a}_1 - 2\mathbf{a}_2 + 4\mathbf{a}_3, \quad \mathbf{a}_1 - \mathbf{a}_2 = 1\mathbf{a}_1 + (-1)\mathbf{a}_2 + 0\mathbf{a}_3 \quad \text{e} \quad 0\mathbf{a}_1 + 0\mathbf{a}_2 + 0\mathbf{a}_3 = \mathbf{0}_7$$

são combinações lineares de $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{a}_3$ (lembre que $\mathbf{0}_7$ é o vetor zero do $\mathbb{R}^7$). A equação (2.2) da seção anterior diz que o vetor $\left[\begin{smallmatrix}14\\-8\end{smallmatrix}\right]$ é a combinação linear dos vetores $\mathbf{x}$, $\mathbf{y}$ e $\mathbf{z}$ com pesos 6, -2 e 4, respectivamente (verifique).

Dizer que um vetor $\mathbf{b}$ de $\mathbb{R}^n$ é uma combinação linear de $\mathbf{a}_1, \ldots, \mathbf{a}_m$ é o mesmo que dizer que $\mathbf{b}$ é **gerado** pelos vetores $\mathbf{a}_1, \ldots, \mathbf{a}_m$. Subentende-se que seja gerado *por uma combinação linear* desses vetores. Na seção 2.5, veremos ainda outras maneiras de dizer isso,[4] e abordaremos o importante problema de determinar quando é que um vetor é uma combinação linear de outros vetores dados.

Exemplo 2.3
Em uma sessão de gravação musical profissional, os elementos de uma música são gravados em “faixas” separadas (vocal, baixo, percussão e piano, por exemplo). A faixa musical completa é obtida pelo processo de “mixagem” (combinação) das faixas individuais. Essa “mixagem”, em sua modalidade mais simples, nada mais é do que uma combinação linear.

Vejamos um exemplo concreto. Na figura 2.4, exibimos quatro faixas de uma gravação da música “Rock and Roll”, do conjunto *Led Zeppelin*: baixo elétrico, bateria, guitarra e vocal.[5] Representamos em cada gráfico, na realidade, um segmento de apenas 40 milissegundos extraído da faixa correspondente.

Estes sinais de áudio[6] podem ser representados por vetores de $\mathbb{R}^n$. De fato, cada faixa da figura 2.4 foi gravada digitalmente a uma taxa de 44.100 amostras por segundo (44,1 kHz). Cada trecho de 40 milissegundos contém, então, $44.100 \times 0{,}040 = 1.764$ amostras. Os sinais ilustrados na figura, portanto, podem ser representados por vetores de $\mathbb{R}^{1764}$, que vamos denotar por $\mathbf{b}$ (**b**aixo), $\mathbf{p}$ (bateria, ou **p**ercussão), $\mathbf{g}$ (**g**uitarra) e $\mathbf{v}$ (**v**ocal).

[4] Essa, talvez, seja uma das principais causas de dificuldade na aprendizagem de álgebra linear: em várias instâncias, existem muitas formas diferentes de dizer a mesma coisa!

[5] Infelizmente, não conseguimos as faixas da gravação original. As gravações apresentadas aqui foram feitas pela “banda-cover” *Boot Led Zeppelin*. (Um trocadilho com “bootleg”, que descreve um artigo pirateado. . .)

[6] O significado de “sinal”, nesse contexto, será dado logo a seguir.

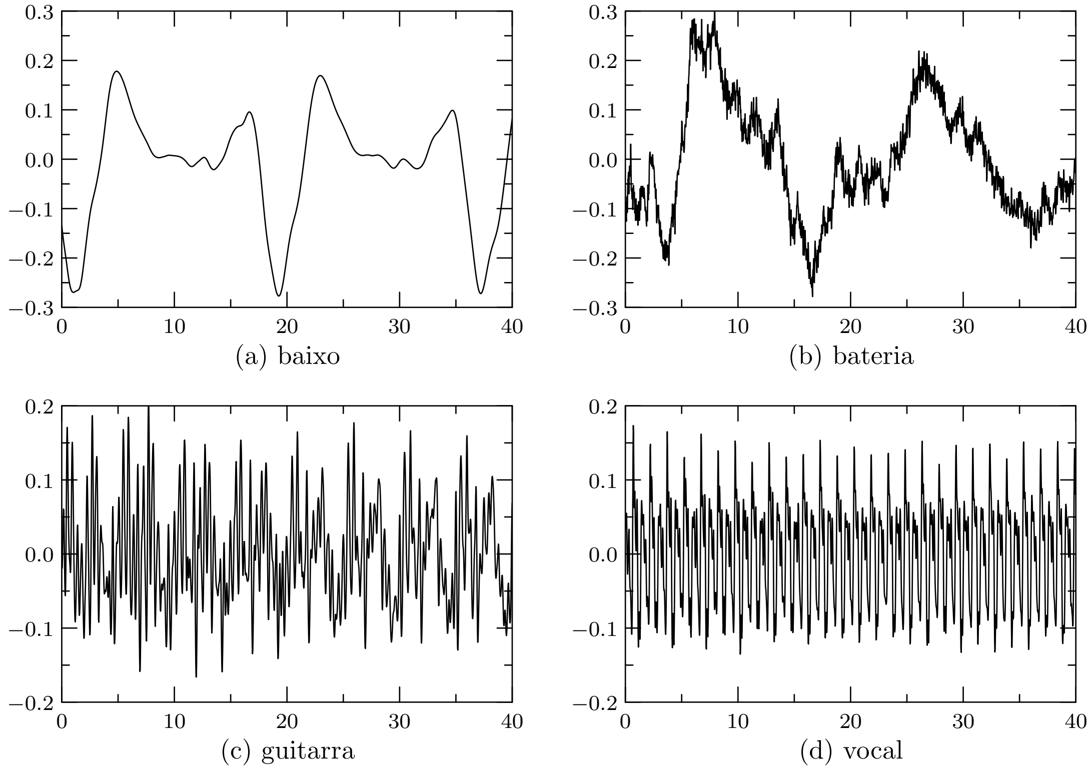
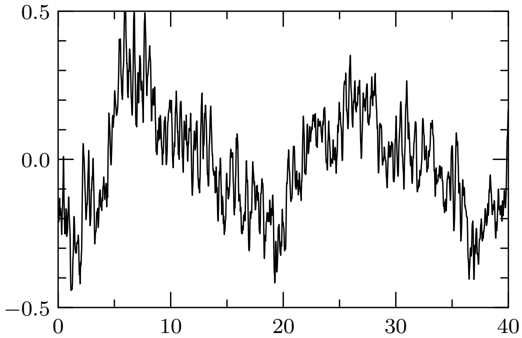


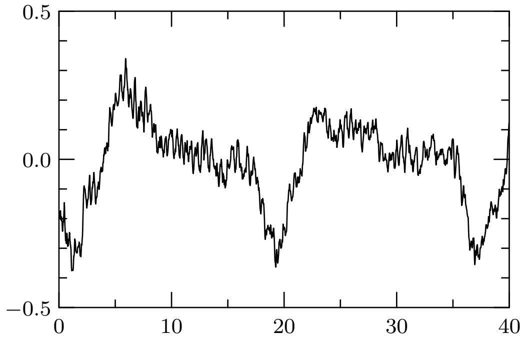
Figura 2.4: Faixas de uma gravação musical. Em cada gráfico, o eixo horizontal representa o tempo em milissegundos.

Podemos obter uma faixa musical completa "mixando" os vetores $\mathbf{b}$, $\mathbf{p}$, $\mathbf{g}$ e $\mathbf{v}$ por simples combinação linear: $c_1\mathbf{b} + c_2\mathbf{p} + c_3\mathbf{g} + c_4\mathbf{v}$. Os pesos c_1 a c_4, neste exemplo, representam os "volumes" de cada faixa. A arte de ajustar tais pesos, de forma que os volumes relativos resultem em uma faixa musical harmoniosa, pertence ao campo da engenharia de som.

Na figura 2.5, exibimos duas combinações lineares (duas "mixagens") das faixas individuais: (a) $\frac{3}{4}\mathbf{b} + \mathbf{p} + \mathbf{g} + \frac{1}{2}\mathbf{v}$, e (b) $\mathbf{b} + \frac{1}{2}\mathbf{p} + \frac{2}{5}\mathbf{g} + \frac{1}{5}\mathbf{v}$. Comparando,



(a) primeira mixagem



(b) segunda mixagem

Figura 2.5: Duas "mixagens" das faixas da figura anterior.

atentamente, os gráficos da figura 2.5 com aqueles da figura 2.4, é possível ver que o "peso" (volume) relativo do baixo $\mathbf{b}$ é, realmente, maior na segunda mixagem do que na primeira. Por outro lado, componentes de alta frequência, características da guitarra $\mathbf{g}$ e do vocal $\mathbf{v}$, são maiores na primeira mixagem.

Admitimos que as "músicas" representadas na figura 2.5 são curtíssimas: elas têm duração de apenas 40 milissegundos! Uma faixa de 4 minutos gravada à taxa de 44,1 kHz teria $44.100 \times 4 \times 60 = 10.584.000$ amostras, isto é, teria de ser representada por um vetor de $\mathbb{R}^{10.584.000}$.

Observação
A álgebra linear é utilizada intensivamente na área de *processamento de sinais*, que lida com a representação e o tratamento matemático de áudio, de imagens, de vídeo e, genericamente, de "mensagens" ou medições de qualquer natureza. Na terminologia dessa área, um "sinal" é uma representação elétrica ou matemática de uma mensagem. Os vetores $\mathbf{b}$, $\mathbf{p}$, $\mathbf{g}$ e $\mathbf{v}$ do exemplo anterior são, portanto, sinais de áudio, bem como qualquer uma de suas combinações lineares.

2.3 O produto de matriz por vetor

Combinações lineares são tão importantes, e serão usadas com tanta frequência, que merecem uma notação mais sucinta do que aquela empregada em (2.3). O *produto de uma matriz por um vetor*, que definiremos abaixo, pode ser usado para esse fim. Mais adiante, veremos outras interpretações importantes para o produto matriz-vetor, e ficará claro que ele é mais do que uma mera conveniência de notação. Por ora, no entanto, esta será sua utilidade principal. Dessa maneira, a frase "uma notação compacta para combinações lineares" pode ser pensada como o subtítulo desta seção.

É conveniente introduzir primeiro uma notação para matrizes que destaque suas colunas. Uma matriz $n \times m$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \tag{2.4}$$

pode ser escrita "por colunas" como

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_m \end{bmatrix}, \tag{2.5}$$

onde $\mathbf{a}_1, \ldots, \mathbf{a}_m$ são os vetores de $\mathbb{R}^n$ dados por

$$\mathbf{a}_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{bmatrix}, \quad \ldots, \quad \mathbf{a}_m = \begin{bmatrix} a_{1m} \\ a_{2m} \\ \vdots \\ a_{nm} \end{bmatrix}. \tag{2.6}$$

É usual chamar os $\mathbf{a}_j$ de "vetores-coluna da matriz A". Compare (2.4) com (2.6) e note que os $\mathbf{a}_j$ são, de fato, as colunas de A.

A matriz

$$\begin{bmatrix} 1 & 2 & -1 \\ -3 & 0 & 5 \end{bmatrix}, \tag{2.7}$$

por exemplo, pode ser escrita como $\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix}$, onde

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \quad \text{e} \quad \mathbf{a}_3 = \begin{bmatrix} -1 \\ 5 \end{bmatrix} \tag{2.8}$$

são seus vetores-coluna.

Agora podemos passar ao objetivo principal desta seção.

Definição 2.4
Seja A uma matriz $n \times m$ com colunas dadas por $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ e seja $\mathbf{x}$ um vetor de $\mathbb{R}^m$ de coordenadas $x_1, x_2, \ldots, x_m$. O **produto de A por $\mathbf{x}$**, denotado por $A\mathbf{x}$, é definido como a combinação linear dos vetores $\mathbf{a}_1, \ldots, \mathbf{a}_m$ com pesos $x_1, \ldots, x_m$, isto é,

$$A\mathbf{x} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_m \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_m\mathbf{a}_m. \tag{2.9}$$

Note que o produto $A\mathbf{x}$ é um vetor de $\mathbb{R}^n$ (onde n é o número de *linhas* de A), e que só está definido quando o número m de componentes de $\mathbf{x}$ é igual ao número de *colunas* de A. Verifique, também, que o produto $A\mathbf{x}$ é nada mais do que uma combinação linear dos vetores-coluna de A (compare (2.3) com (2.9)).

Exemplo 2.5

Vamos calcular o produto da matriz $\begin{bmatrix} 1 & 2 & -1 \\ -3 & 0 & 5 \end{bmatrix}$ de (2.7) pelo vetor $\begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix}$. Usando a definição acima, temos

$$\begin{bmatrix} 1 & 2 & -1 \\ -3 & 0 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ -3 \end{bmatrix} + 3 \begin{bmatrix} 2 \\ 0 \end{bmatrix} + (-1) \begin{bmatrix} -1 \\ 5 \end{bmatrix} =$$
$$= \begin{bmatrix} 2 \\ -6 \end{bmatrix} + \begin{bmatrix} 6 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ -5 \end{bmatrix} = \begin{bmatrix} 9 \\ -11 \end{bmatrix}.$$

É provável que você já tenha estudado o produto matriz-vetor no ensino médio. A definição 2.4 talvez não se pareça muito com o produto que você já conhece. Mas não fique confuso, pois a expressão (2.9) dada acima coincide com o produto "habitual". Vamos verificar isso, escrevendo $A\mathbf{x}$ "por extenso". Se A é a matriz dada em (2.4), então

$$A\mathbf{x} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{bmatrix} + \cdots + x_m \begin{bmatrix} a_{1m} \\ a_{2m} \\ \vdots \\ a_{nm} \end{bmatrix}$$
$$= \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m \end{bmatrix}. \tag{2.10}$$

A primeira linha acima é simplesmente (2.9) reescrita (verifique!) e a última igualdade segue diretamente da aplicação das operações vetoriais (veja a subseção 2.1.5). Verifique, agora, que o vetor à direita, em (2.10), é precisamente o produto $A\mathbf{x}$ "habitual".

Proposição 2.6 (Propriedades algébricas do produto matriz-vetor)
Se A é uma matriz $n \times m$ qualquer, $\mathbf{u}$ e $\mathbf{v}$ são vetores de $\mathbb{R}^m$, e α é um escalar, então valem as propriedades abaixo.

(a) $A(\mathbf{u}+\mathbf{v}) = A\mathbf{u} + A\mathbf{v}$ (b) $A(\alpha\mathbf{u}) = \alpha A\mathbf{u}$

É fácil deduzir essas propriedades. Vamos verificar apenas (a), e deixamos a verificação de (b) para o leitor (veja o exercício P2.8). Se $\mathbf{a}_1, \ldots, \mathbf{a}_m$ são as colunas da matriz A, e $u_1, \ldots, u_m$ e $v_1, \ldots, v_m$ são, respectivamente, as coordenadas de $\mathbf{u}$ e $\mathbf{v}$, então

$$\begin{aligned} A(\mathbf{u}+\mathbf{v}) &= (u_1+v_1)\mathbf{a}_1 + (u_2+v_2)\mathbf{a}_2 + \cdots + (u_m+v_m)\mathbf{a}_m \\ &= (u_1\mathbf{a}_1 + u_2\mathbf{a}_2 + \cdots + u_m\mathbf{a}_m) + (v_1\mathbf{a}_1 + v_2\mathbf{a}_2 + \cdots + v_m\mathbf{a}_m) \\ &= A\mathbf{u} + A\mathbf{v}. \end{aligned}$$

A primeira igualdade acima decorre diretamente da definição 2.4 (as coordenadas de $\mathbf{u}+\mathbf{v}$ são $u_1+v_1, \ldots, u_m+v_m$); a segunda, das propriedades algébricas das operações vetoriais; e a terceira, novamente, da definição 2.4.

Aplicações repetidas dessas propriedades permitem mostrar o seguinte resultado. Se A é uma matriz $n \times m$, $\mathbf{v}_1, \ldots, \mathbf{v}_q$ são vetores de $\mathbb{R}^m$ e $c_1, \ldots, c_q$ são escalares, então

$$A\big(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_q\mathbf{v}_q\big) = c_1A\mathbf{v}_1 + c_2A\mathbf{v}_2 + \cdots + c_qA\mathbf{v}_q. \tag{2.11}$$

Note que a expressão dentro dos parênteses, à esquerda, é uma combinação linear dos vetores $\mathbf{v}_1, \ldots, \mathbf{v}_q$. A expressão à direita, por sua vez, é uma combinação linear dos vetores $A\mathbf{v}_1, \ldots, A\mathbf{v}_q$. A "moral" da equação (2.11) é dizer que o produto matriz-vetor exibe "bom comportamento" com respeito a combinações lineares, no seguinte sentido: o produto por uma combinação linear é a combinação linear dos produtos, com os mesmos pesos.

As propriedades discutidas acima terão papel crucial no capítulo 5.

2.4 Notação vetorial para sistemas lineares

Os conceitos que vimos acima podem ser usados para escrever sistemas lineares de forma mais simples e sucinta do que fizemos no capítulo 1. Desejamos chamar a atenção da leitora, ou do leitor, para este fato. Pense nesta seção como um aparte que, de certa maneira, não pertence a este capítulo, pois não traz nenhum conceito novo sobre vetores. A notação que veremos aqui, no entanto, será útil para as próximas seções, por isso vamos introduzi-la agora.

Vamos começar com um exemplo. O sistema de equações lineares

$$\begin{cases} x_1 + 2x_2 - x_3 = 9 \\ -3x_1 + 5x_3 = -11 \end{cases} \tag{2.12}$$

pode ser escrito, de acordo com (2.1) (página 36), na forma vetorial

$$\begin{bmatrix} x_1 + 2x_2 - x_3 \\ -3x_1 + 5x_3 \end{bmatrix} = \begin{bmatrix} 9 \\ -11 \end{bmatrix}. \tag{2.13}$$

Usando as operações definidas na seção 2.1.5, podemos reescrever (2.13) novamente como

$$x_1\begin{bmatrix} 1 \\ -3 \end{bmatrix} + x_2\begin{bmatrix} 2 \\ 0 \end{bmatrix} + x_3\begin{bmatrix} -1 \\ 5 \end{bmatrix} = \begin{bmatrix} 9 \\ -11 \end{bmatrix}. \tag{2.14}$$

Representando a lista das variáveis x_1, x_2 e x_3 como um vetor $\mathbf{x}$ de $\mathbb{R}^3$, e usando a definição do produto matriz-vetor, esta equação pode ser reescrita ainda como

$$\begin{bmatrix} 1 & 2 & -1 \\ -3 & 0 & 5 \end{bmatrix}\mathbf{x} = \begin{bmatrix} 9 \\ -11 \end{bmatrix}. \tag{2.15}$$

A equação acima nada mais é do que um sistema linear. Enfatizamos que (2.13), (2.14) e (2.15) são, de fato, formas diferentes de escrever o sistema (2.12). Observe que é fácil "ler", diretamente de (2.15), a matriz completa desse sistema, que é

$$\left[\begin{array}{rrr|r} 1 & 2 & -1 & 9 \\ -3 & 0 & 5 & -11 \end{array}\right].$$

Agora tratemos do caso geral. Um sistema qualquer de n equações lineares envolvendo m variáveis

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m = b_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m = b_n \end{cases} \tag{2.16}$$

pode ser reescrito na forma vetorial

$$x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{bmatrix} + \cdots + x_m \begin{bmatrix} a_{1m} \\ a_{2m} \\ \vdots \\ a_{nm} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}. \tag{2.17}$$

A equação acima é igual a

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_m\mathbf{a}_m = \mathbf{b}, \tag{2.18}$$

onde os vetores $\mathbf{a}_1$, $\mathbf{a}_2$, ... , $\mathbf{a}_m$ e $\mathbf{b}$ são dados por

$$\mathbf{a}_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{bmatrix}, \quad \dots, \quad \mathbf{a}_m = \begin{bmatrix} a_{1m} \\ a_{2m} \\ \vdots \\ a_{nm} \end{bmatrix} \quad \text{e} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

Observe que esses são vetores de $\mathbb{R}^n$, e n é o número de equações do sistema (2.16).

Representando a lista das variáveis x_1, x_2, ... , x_m por um vetor $\mathbf{x}$ de $\mathbb{R}^m$, e usando o produto matriz-vetor, a equação (2.18) pode ser reescrita na forma

$$\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_m \end{bmatrix} \mathbf{x} = \mathbf{b}. \tag{2.19}$$

Chamando a matriz $\begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_m \end{bmatrix}$ de A, como em (2.5), (2.19) torna-se

$$A\mathbf{x} = \mathbf{b}. \tag{2.20}$$

Repare que A é precisamente a matriz de *coeficientes* do sistema (2.16) (compare (2.16) com (2.4) da página 43). O vetor $\mathbf{b}$ é o *vetor dos termos independentes* do sistema. A matriz *completa* do sistema é escrita, usando a notação "por colunas", como

$$\left[\begin{array}{cccc|c} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_m & \mathbf{b} \end{array}\right].$$

Podemos escrever essa matriz de forma ainda mais compacta como $\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]$.

Enfatizamos, novamente, que (2.16), (2.17), (2.18), (2.19) e (2.20) são formas diferentes (ordenadas da mais prolixa para a mais sucinta) de escrever *o mesmo sistema linear*. Você deve se habituar com cada uma destas notações, pois elas serão muito usadas. Ao deparar-se com algo como (2.18) ou (2.20), você deve perceber a correspondência com (2.16) imediatamente, sem precisar pensar muito!

A notação compacta $A\mathbf{x} = \mathbf{b}$ de (2.20) será particularmente útil. Lembre-se de que, no contexto de sistemas lineares, a matriz de coeficientes A e o vetor $\mathbf{b}$ tipicamente são dados, e $\mathbf{x}$ é o vetor cujas coordenadas são as variáveis ou "incógnitas" $x_1, \ldots, x_m$. Você pode pensar em $\mathbf{x}$ como a "incógnita vetorial" da equação $A\mathbf{x} = \mathbf{b}$.

Podemos, então, reformular o conceito de *solução* de um sistema: um vetor $\mathbf{u}$ é dito uma **solução** do sistema linear $A\mathbf{x} = \mathbf{b}$ se $A\mathbf{u} = \mathbf{b}$, ou seja, se a igualdade vetorial torna-se verdadeira quando substituímos $\mathbf{x}$ por $\mathbf{u}$. O **conjunto-solução** de $A\mathbf{x} = \mathbf{b}$ é o conjunto de todos os vetores $\mathbf{u}$ de $\mathbb{R}^m$ tais que $A\mathbf{u} = \mathbf{b}$.

A notação vetorial é conveniente, também, para *descrever* tais conjuntos-solução. Revisitemos alguns exemplos do capítulo 1, para ver como isso é feito.

Exemplo 2.7
Considere o sistema (1.20) do exemplo 1.9, na página 22. Uma descrição paramétrica de seu conjunto-solução é dada por (1.22):

$$\begin{cases} x_1 = 5 + 3x_3, \\ x_2 = -3 - 2x_3, \\ x_3 \text{ é uma variável livre,} \\ x_4 = 0. \end{cases}$$

Usando a notação vetorial, podemos reescrever essa descrição como

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 5 + 3x_3 \\ -3 - 2x_3 \\ x_3 \\ 0 \end{bmatrix} \qquad (x_3 \text{ é uma variável livre}). \tag{2.21}$$

Note que $\mathbf{x}$ é um vetor de $\mathbb{R}^4$, pois há quatro variáveis no sistema (1.20). Perceba, também, que a descrição (2.21) não impõe nenhuma restrição genuína sobre x_3, já que essa é uma variável livre. As igualdades entre as terceiras coordenadas em (2.21) dizem, simplesmente, "$x_3 = x_3$".

Podemos reescrever (2.21) em uma forma que será ainda mais conveniente. Para isso, colocamos "em evidência" a variável livre x_3:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 5 + 3x_3 \\ -3 - 2x_3 \\ x_3 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 3 \\ -2 \\ 1 \\ 0 \end{bmatrix}. \tag{2.22}$$

Repare que a última igualdade acima é válida para qualquer valor de x_3. Basta verificá-la coordenada por coordenada.

Dizemos que (2.22) é uma **descrição vetorial paramétrica** do conjunto-solução do sistema (1.20). Vejamos mais um exemplo.

Exemplo 2.8
Considere, agora, o sistema (1.16) do exemplo 1.7, na página 19. Obtivemos uma descrição paramétrica (não-vetorial) de seu conjunto-solução em (1.18):

$$\begin{cases} x_1 = 3 - 2x_2 - 5x_4, \\ x_2 \text{ é uma variável livre}, \\ x_3 = 7 + 3x_4, \\ x_4 \text{ é uma variável livre.} \end{cases}$$

Para obter essa descrição na forma vetorial, procedemos como no exemplo anterior. Escrevemos as variáveis x_j como coordenadas de um vetor $\mathbf{x}$, usamos as igualdades acima, e, depois, colocamos "em evidência" as variáveis livres:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 3 - 2x_2 - 5x_4 \\ x_2 \\ 7 + 3x_4 \\ x_4 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 7 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -5 \\ 0 \\ 3 \\ 1 \end{bmatrix}.$$

Perceba, novamente, que, com relação às variáveis livres, a descrição acima diz apenas "$x_2 = x_2$" e "$x_4 = x_4$".

Exemplo 2.9
Por fim, consideramos o sistema (1.23) do exemplo 1.10, na página 23. A descrição paramétrica de seu conjunto-solução é (1.25):

$$\begin{cases} x_1 = 24, \\ x_2 = 7, \\ x_3 = 4. \end{cases}$$

A descrição *vetorial* paramétrica, então, é

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 24 \\ 7 \\ 4 \end{bmatrix}.$$

Como observamos no exemplo 1.10, a terminologia, nesse caso, é artificiosa, pois não há "parâmetro" algum nas descrições acima. Isso é porque o sistema (1.23) não possui variáveis livres.

Daqui para diante, *resolver* um sistema linear significará obter uma descrição *vetorial* de seu conjunto-solução, ou determinar que ele é impossível. Reveja o procedimento descrito na subseção 1.7.3 e acrescente um sétimo passo: "obtenha uma descrição *vetorial* paramétrica do conjunto-solução". Com prática, você será capaz de ir diretamente do passo 4 a esse sétimo passo, pulando o 5 e o 6. Recomendamos os exercícios P2.9 e P2.10.

2.5 O espaço gerado por vetores (o *span*)

Nesta seção, voltamos a colocar o conceito de combinação linear em primeiro plano. Uma rápida revisão da definição 2.2 é aconselhável (ver página 40).

Dado um conjunto de vetores de $\mathbb{R}^n$, podemos considerar o conjunto de *todas* as suas combinações lineares (conforme a definição a seguir). O resultado é um tipo especial de subconjunto de $\mathbb{R}^n$, chamado *subespaço*, que tem grande importância em álgebra linear. Estudaremos subespaços de $\mathbb{R}^n$ no capítulo 4, mas, até lá, vamos usar essa terminologia sem muita justificativa. Por ora, considere que um subespaço é meramente um subconjunto de $\mathbb{R}^n$, com certas propriedades especiais que serão discutidas mais adiante.

Definição 2.10
Dados os vetores $\mathbf{a}_1$, $\mathbf{a}_2$, ... , $\mathbf{a}_m$ de $\mathbb{R}^n$, *o conjunto de todas as suas combinações lineares* é chamado de **subespaço gerado por** $\mathbf{a}_1$, $\mathbf{a}_2$, ... , $\mathbf{a}_m$, e é denotado por $\operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$. Em outras palavras, $\operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ é o conjunto de todos os vetores de $\mathbb{R}^n$ que podem ser escritos na forma

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_m\mathbf{a}_m,$$

onde x_1, x_2, ... , x_m são escalares (pesos) quaisquer.

Enfatizamos que $\operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ é um subconjunto de $\mathbb{R}^n$, onde n é o número de coordenadas dos vetores $\mathbf{a}_1$, ... , $\mathbf{a}_m$. Não confunda n com m!

Observação
A expressão “o subespaço gerado por $\mathbf{a}_1$, ... , $\mathbf{a}_m$” traduz-se para o inglês como “*the span of* $\mathbf{a}_1$, ... , $\mathbf{a}_m$” (daí a notação $\operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$). O substantivo *span*, em inglês corrente, significa algo como “região (ou período) de abrangência”, “alcance” ou “alçada”. Assim, dizer que $\mathbf{b}$ está no *span* de $\mathbf{a}_1$, ... , $\mathbf{a}_m$ é dizer que $\mathbf{b}$ está ao alcance (via combinações lineares) dos vetores $\mathbf{a}_1$, ... , $\mathbf{a}_m$. Dizer que $\mathbf{b}$ *não* pertence ao *span* de $\mathbf{a}_1$, ... , $\mathbf{a}_m$ é como dizer que $\mathbf{b}$ está “fora da alçada” desses vetores. Continuaremos a cometer esses anglicismos ocasionalmente, e faremo-lo sem remorso, porque o significado corrente de *span* nos ajuda a ilustrar o conceito matemático.

É fácil descrever geometricamente subespaços gerados por vetores de $\mathbb{R}^2$ ou de $\mathbb{R}^3$, como mostra o exemplo a seguir. Os exercícios P2.16, P2.17, P2.21 e P2.22 contêm mais exemplos dessa natureza (especialmente o último).

Exemplo 2.11
Considere os vetores $\mathbf{u}$ e $\mathbf{v}$ de $\mathbb{R}^3$ ilustrados na figura 2.6(a). O *span* de $\mathbf{u}$ e $\mathbf{v}$ é precisamente o plano que contém ambos os vetores. Em outras palavras, qualquer vetor desse plano pode ser escrito como uma combinação linear de $\mathbf{u}$ e $\mathbf{v}$, mas vetores que não se encontram nesse plano estão “fora do alcance” de $\mathbf{u}$ e $\mathbf{v}$, isto é, não são combinações lineares desses dois vetores.

Os vetores $\mathbf{u}'$ e $\mathbf{v}'$ de $\mathbb{R}^3$, ilustrados na figura 2.6(b), por outro lado, são *colineares* (ou seja, estão ambos contidos em uma mesma reta) e não-nulos. Isso implica que qualquer combinação linear de tais vetores é simplesmente um múltiplo de $\mathbf{u}'$ (ou de $\mathbf{v}'$), conforme o exercício P2.20. O *span* de $\mathbf{u}'$ e $\mathbf{v}'$, portanto, é justamente a reta que contém esses vetores.

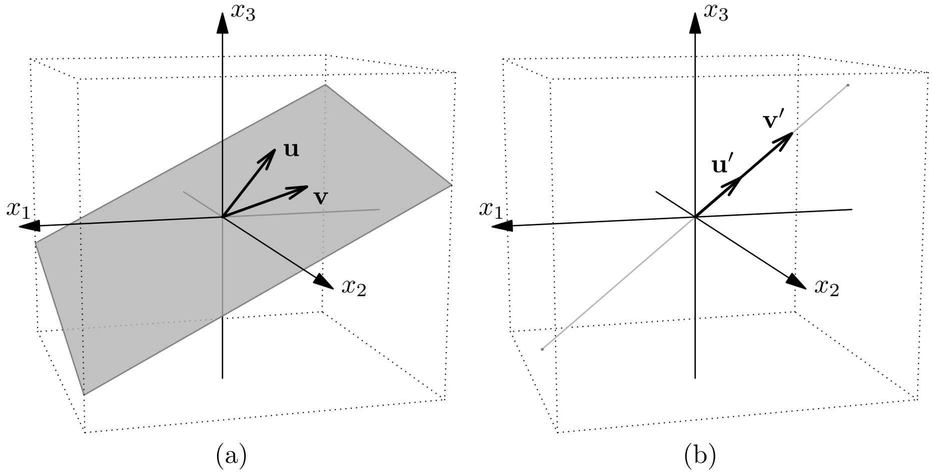


Figura 2.6: Span$\{\mathbf{u}, \mathbf{v}\}$ é o plano hachurado em (a). Já Span$\{\mathbf{u}', \mathbf{v}'\}$ é a reta ilustrada em (b).

Determinar se um vetor $\mathbf{b} \in \mathbb{R}^n$ *pertence a* Span$\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ *é o mesmo que determinar se* $\mathbf{b}$ *é uma combinação linear dos vetores* $\mathbf{a}_1, \ldots, \mathbf{a}_m$.[7]

Vejamos como se faz essa determinação “na prática”.

Exemplo 2.12
Considere os vetores de $\mathbb{R}^3$ dados por

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \quad \text{e} \quad \mathbf{b} = \begin{bmatrix} 2 \\ 3 \\ -5 \end{bmatrix}. \tag{2.23}$$

Determine se $\mathbf{b}$ pertence a Span$\{\mathbf{a}_1, \mathbf{a}_2\}$.

Solução: Temos que determinar se $\mathbf{b}$ é uma combinação linear de $\mathbf{a}_1$ e $\mathbf{a}_2$, ou seja, se existem escalares (pesos) x_1 e x_2 tais que $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = \mathbf{b}$. Escrevendo essa equação vetorial “por extenso”, obtemos

$$x_1 \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ -5 \end{bmatrix}. \tag{2.24}$$

Portanto, $\mathbf{b}$ *é uma combinação linear de* $\mathbf{a}_1$ *e* $\mathbf{a}_2$ *se e somente se a equação* (2.24) *tem solução*. Como foi discutido na seção anterior, essa equação corresponde exatamente ao sistema linear

$$\begin{cases} x_1 + x_2 = 2 \\ 2x_1 + 3x_2 = 3 \\ -x_1 + 2x_2 = -5. \end{cases}$$

[7] Esta afirmativa é trivial, pois Span$\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ é *definido* como o conjunto das combinações lineares de $\mathbf{a}_1, \ldots, \mathbf{a}_m$ (definição 2.10). Seria como dizer: “Determinar se o ornitorrinco pertence ao conjunto dos mamíferos é o mesmo que determinar se o ornitorrinco é um mamífero.”
O ornitorrinco, a propósito, é um mamífero.

Determinar se um sistema é possível é um problema que já conhecemos (ver as subseções 1.7.2 e 1.8.1, em particular as proposições 1.11 e 1.12). Escalonando a matriz completa do sistema (2.24), obtemos

$$\begin{bmatrix} \boxed{1} & 1 & 2 \\ 2 & 3 & 3 \\ -1 & 2 & -5 \end{bmatrix} \xrightarrow[\ell_3 \to \ell_3 + \ell_1]{\ell_2 \to \ell_2 - 2\ell_1} \begin{bmatrix} \boxed{1} & 1 & 2 \\ 0 & \boxed{1} & -1 \\ 0 & 3 & -3 \end{bmatrix} \xrightarrow{\ell_3 \to \ell_3 - 3\ell_2} \begin{bmatrix} \boxed{1} & 1 & 2 \\ 0 & \boxed{1} & -1 \\ 0 & 0 & 0 \end{bmatrix}. \tag{2.25}$$

Como não há linhas do tipo (1.26) na matriz escalonada obtida (não há equações do tipo $0 = \blacksquare$ no sistema associado), o sistema (2.24) é possível. Portanto, $\mathbf{b}$ é uma combinação linear dos vetores $\mathbf{a}_1$ e $\mathbf{a}_2$. Com isso, chegamos ao fim do exercício e à seguinte conclusão: o vetor $\mathbf{b}$ pertence a $\mathrm{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$. □

Observe que não foi necessário resolver o sistema (2.24) completamente para responder à questão proposta. Em particular, não foi necessário obter a forma escalonada reduzida em (2.25). No entanto, se quisermos escrever $\mathbf{b}$ explicitamente como combinação linear de $\mathbf{a}_1$ e $\mathbf{a}_2$, então, aí sim, será necessário resolver o sistema (2.24). Verifique que a (única) solução do sistema (2.24) é $x_1 = 3$, $x_2 = -1$, e, portanto, vale $\mathbf{b} = 3\mathbf{a}_1 - \mathbf{a}_2$.

Exemplo 2.13
Determine se o vetor $\mathbf{d} = \left[\begin{smallmatrix} 2 \\ 2 \\ -5 \end{smallmatrix}\right]$ pertence a $\mathrm{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$, onde $\mathbf{a}_1$ e $\mathbf{a}_2$ são os vetores de $\mathbb{R}^3$ do exemplo anterior.

Solução: Procedemos como no caso anterior. Queremos determinar se o sistema $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = \mathbf{d}$ é possível. Para isso, escalonamos a sua matriz completa:

$$\begin{bmatrix} \boxed{1} & 1 & 2 \\ 2 & 3 & 2 \\ -1 & 2 & -5 \end{bmatrix} \xrightarrow[\ell_3 \to \ell_3 + \ell_1]{\ell_2 \to \ell_2 - 2\ell_1} \begin{bmatrix} \boxed{1} & 1 & 2 \\ 0 & \boxed{1} & -2 \\ 0 & 3 & -3 \end{bmatrix} \xrightarrow{\ell_3 \to \ell_3 - 3\ell_2} \begin{bmatrix} \boxed{1} & 1 & 2 \\ 0 & \boxed{1} & -2 \\ 0 & 0 & \boxed{3} \end{bmatrix}$$

O sistema $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = \mathbf{d}$ é *impossível*, pois é equivalente a um sistema que tem a equação $0 = 3$, como se pode ver acima. Portanto, $\mathbf{d}$ não é uma combinação linear de $\mathbf{a}_1$ e $\mathbf{a}_2$, isto é, $\mathbf{d}$ *não pertence a* $\mathrm{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$. □

Observação 2.14
O vetor zero de $\mathbb{R}^n$ *sempre* pertence a $\mathrm{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$, quaisquer que sejam os vetores $\mathbf{a}_1, \ldots, \mathbf{a}_m$ de $\mathbb{R}^n$, pois

$$\mathbf{0}_n = 0\mathbf{a}_1 + 0\mathbf{a}_2 + \cdots + 0\mathbf{a}_m.$$

Ou seja, o vetor zero é sempre uma combinação linear de $\mathbf{a}_1, \ldots, \mathbf{a}_m$.

Observação 2.15
Cada um dos vetores $\mathbf{a}_1, \ldots, \mathbf{a}_m$ também *sempre* pertence a $\mathrm{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$. Para verificar que $\mathbf{a}_1 \in \mathrm{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$, por exemplo, basta observar que $\mathbf{a}_1 = 1\mathbf{a}_1 + 0\mathbf{a}_2 + \cdots + 0\mathbf{a}_m$. A verificação para os outros vetores é análoga.

Existem diversas maneiras de se dizer que um vetor $\mathbf{b}$ de $\mathbb{R}^n$ pertence ao subespaço gerado por $\mathbf{a}_1, \ldots, \mathbf{a}_m$. Para a sua conveniência, reunimos as mais usuais na proposição a seguir.[8]

Proposição 2.16
As seguintes afirmativas são equivalentes (isto é, ou são todas verdadeiras, ou então são todas falsas):

(a) $\mathbf{b} \in \operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$.[9]

(b) $\mathbf{b}$ *é gerado por* $\mathbf{a}_1, \ldots, \mathbf{a}_m$.[10]

(c) $\mathbf{b}$ *é uma combinação linear dos vetores* $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$.

(d) *Existe ao menos uma lista de pesos* $x_1, \ldots, x_m$ *tal que* $\mathbf{b} = x_1\mathbf{a}_1 + \cdots + x_m\mathbf{a}_m$.

(e) *O sistema linear* $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_m\mathbf{a}_m = \mathbf{b}$ *é possível.*

Esperamos que a equivalência entre as afirmativas esteja clara, à luz das definições e dos exemplos acima. Geralmente, a afirmativa (e) é usada "na prática" para determinar se as outras valem ou não (como foi feito nos exemplos 2.12 e 2.13).

2.6 O espaço-coluna de uma matriz

A definição a seguir está intimamente relacionada à definição 2.10.

Definição 2.17
Seja $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_m \end{bmatrix}$ uma matriz $n \times m$. O **espaço das colunas** de A é o subespaço gerado pelos vetores $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ de $\mathbb{R}^n$. Denotamos o espaço das colunas de A por $\operatorname{Col} A$.

Em símbolos, a definição diz que $\operatorname{Col} A = \operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$, onde $\mathbf{a}_1, \ldots, \mathbf{a}_m$ são os vetores-coluna de A. O espaço das colunas também é chamado de **espaço gerado pelas colunas** ou, simplesmente, de **espaço-coluna** de A.

Note que o espaço-coluna de A é um subconjunto de $\mathbb{R}^n$, onde n é o número de *linhas* da matriz A. O conjunto $\operatorname{Col} A$ é, de fato, um *subespaço* de $\mathbb{R}^n$, como veremos no capítulo 4.

Exemplo 2.18
Sejam

$$A = \begin{bmatrix} 1 & 1 \\ 2 & 3 \\ -1 & 2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2 \\ 3 \\ -5 \end{bmatrix} \quad \text{e} \quad \mathbf{d} = \begin{bmatrix} 2 \\ 2 \\ -5 \end{bmatrix}.$$

Determine se $\mathbf{b}$ e $\mathbf{d}$ pertencem a $\operatorname{Col} A$.

[8] Admitimos que há *muita* redundância no enunciado da proposição. Nossa intenção é ajudar o leitor a assimilar os conceitos e abordar os exercícios. Pecamos pela prolixidade, mas não pela falta de clareza (ou assim esperamos).

[9] Lembre que o símbolo "$\in$" significa "pertence a".

[10] Introduzimos essa terminologia na página 40.

Solução: Isso é uma mera repetição dos exemplos 2.12 e 2.13. De fato, por definição $\operatorname{Col} A = \operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$, onde $\mathbf{a}_1$ e $\mathbf{a}_2$ são os vetores-coluna de A, dados em (2.23), na página 50. Já sabemos que $\mathbf{b} \in \operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$ (exemplo 2.12) e que $\mathbf{d} \notin \operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$ (exemplo 2.13), portanto, $\mathbf{b} \in \operatorname{Col} A$ e $\mathbf{d} \notin \operatorname{Col} A$. □

A questão está solucionada, mas vamos explorar esse exemplo um pouco mais. Vimos, no exemplo 2.12, que $\mathbf{b} \in \operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$, já que o sistema $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = \mathbf{b}$ é possível. Esse sistema pode ser escrito na forma compacta $A\mathbf{x} = \mathbf{b}$ (ver seção 2.4). Similarmente, o sistema $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = \mathbf{d}$ pode ser escrito como $A\mathbf{x} = \mathbf{d}$. No exemplo 2.13, vimos que esse sistema é impossível, logo $\mathbf{d} \notin \operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$.

Em síntese, $\mathbf{b} \in \operatorname{Col} A = \operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$, pois o sistema $A\mathbf{x} = \mathbf{b}$ é *possível*. Já $\mathbf{d} \notin \operatorname{Col} A$, pois o sistema $A\mathbf{x} = \mathbf{d}$ é *impossível*.

Generalizando o exemplo acima, temos o resultado a seguir.

Proposição 2.19
Seja A uma matriz $n \times m$ qualquer. Um vetor $\mathbf{b}$ de $\mathbb{R}^n$ pertence a $\operatorname{Col} A$ se e somente se o sistema linear $A\mathbf{x} = \mathbf{b}$ é possível.

De fato, se $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_m \end{bmatrix}$, então $\operatorname{Col} A = \operatorname{Span}\{\mathbf{a}_1, \dots, \mathbf{a}_m\}$. Dizer que $\mathbf{b} \in \operatorname{Span}\{\mathbf{a}_1, \dots, \mathbf{a}_m\}$ equivale a dizer que o sistema $x_1\mathbf{a}_1 + \cdots + x_m\mathbf{a}_m = \mathbf{b}$ é possível (ver afirmativas (a) e (e) na proposição 2.16). Esse sistema, por sua vez, é exatamente o sistema $A\mathbf{x} = \mathbf{b}$.

Podemos, assim, estender a proposição 2.16, acrescentando mais algumas afirmativas equivalentes. Quando $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_m \end{bmatrix}$, estas são "formas compactas" de escrever as afirmativas (a), (b), (c) e (e), respectivamente:

(a′) $\mathbf{b} \in \operatorname{Col} A$.

(b′) $\mathbf{b}$ é gerado pelas colunas de A.

(c′) $\mathbf{b}$ é uma combinação linear das colunas de A.

(e′) O sistema linear $A\mathbf{x} = \mathbf{b}$ é possível.

2.7 Conjuntos que geram $\mathbb{R}^n$

Dados os vetores $\mathbf{a}_1$, $\mathbf{a}_2$, … , $\mathbf{a}_m$ e $\mathbf{y}$ de $\mathbb{R}^n$, já sabemos abordar o problema: "o vetor $\mathbf{y}$ é gerado por $\mathbf{a}_1$, … , $\mathbf{a}_m$?" A chave está na afirmativa (e) da proposição 2.16: basta determinar se o sistema

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_m\mathbf{a}_m = \mathbf{y} \tag{2.26}$$

é possível. Uma questão mais interessante é determinar se os vetores $\mathbf{a}_1$, … , $\mathbf{a}_m$ são capazes de gerar *qualquer vetor* de $\mathbb{R}^n$.

Definição 2.20
Se o sistema (2.26) tem solução, *qualquer que seja* o vetor $\mathbf{y}$ de $\mathbb{R}^n$, dizemos que **o conjunto** $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m\}$ **gera o** $\mathbb{R}^n$. Podemos também dizer, mais informalmente, que os vetores $\mathbf{a}_1$, … , $\mathbf{a}_m$ geram o $\mathbb{R}^n$.

Caracterizar os conjuntos que geram $\mathbb{R}^n$ é uma questão importante em álgebra linear. Antes de abordar essa questão no contexto geral, vamos considerar dois exemplos.

Exemplo 2.21
Sejam

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \quad \text{e} \quad \mathbf{a}_3 = \begin{bmatrix} -1 \\ 1 \\ 10 \end{bmatrix}.$$

Vamos investigar se o conjunto $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ gera o $\mathbb{R}^3$.

Começamos reformulando essa questão. O que queremos é saber se qualquer vetor $\mathbf{y}$ de $\mathbb{R}^3$ pode ser gerado por $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{a}_3$, isto é, se o sistema

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 = \mathbf{y} \tag{2.27}$$

é possível para *qualquer escolha* de $\mathbf{y}$.

Vamos denotar as coordenadas de $\mathbf{y}$ por y_1, y_2 e y_3, e abordar essa questão exatamente como nos exemplos 2.12 e 2.13. Assim, vamos escalonar a matriz completa do sistema (2.27):

$$\left[\begin{array}{ccc|c} \boxed{1} & 1 & -1 & y_1 \\ 2 & 3 & 1 & y_2 \\ -1 & 2 & 10 & y_3 \end{array}\right] \xrightarrow[\ell_3 \to \ell_3 + \ell_1]{\ell_2 \to \ell_2 - 2\ell_1} \left[\begin{array}{ccc|c} \boxed{1} & 1 & -1 & y_1 \\ 0 & \boxed{1} & 3 & y_2 - 2y_1 \\ 0 & 3 & 9 & y_3 + y_1 \end{array}\right] \longrightarrow$$

$$\xrightarrow{\ell_3 \to \ell_3 - 3\ell_2} \left[\begin{array}{ccc|c} \boxed{1} & 1 & -1 & y_1 \\ 0 & \boxed{1} & 3 & -2y_1 + y_2 \\ 0 & 0 & 0 & 7y_1 - 3y_2 + y_3 \end{array}\right] \tag{2.28}$$

A posição hachurada é a chave de nossa investigação. Perceba que essa pode ou não ser uma posição-pivô da matriz completa, dependendo do valor de $7y_1 - 3y_2 + y_3$. Se esse valor for *igual a zero*, então a última coluna da matriz completa será não-pivô, e o sistema (2.27) será *possível.* Por outro lado, se o valor for *diferente de zero*, então a terceira linha da matriz escalonada obtida acima representará uma equação do tipo $0 = \blacksquare$. Em outras palavras, a última coluna da matriz completa será uma coluna-pivô, e, nesse caso, o sistema (2.27) será *impossível.*

O conjunto $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$, portanto, *não gera* o $\mathbb{R}^3$, pois, como acabamos de ver, *não* é para qualquer vetor $\mathbf{y}$ que o sistema (2.27) é possível. Ou seja, existem vetores de $\mathbb{R}^3$ que *não* são gerados por $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{a}_3$. O vetor $\mathbf{y}' = \left[\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix}\right]$ é um exemplo (há muitos outros), pois, nesse caso, $7y_1 - 3y_2 + y_3 = 7 \neq 0$. Por outro lado, o vetor $\mathbf{y}'' = \left[\begin{smallmatrix} 0 \\ 1 \\ 3 \end{smallmatrix}\right]$ *é* gerado por $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{a}_3$ (verifique!). A descrição do conjunto dos vetores que são gerados por $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{a}_3$ é o objetivo do exercício P2.21.

Cuidado!
Frisamos que a chave na discussão acima é o valor de $7y_1 - 3y_2 + y_3$, que ocupa a posição "talvez pivô, talvez não" na matriz *escalonada* em (2.28). O valor de y_3, que ocupa esta posição na matriz completa *original*, tem pouca relevância nesta

discussão. Ou seja, o valor de y_3, por si só, não determina se o sistema (2.27) é ou não possível. O fato de que y_3 ocupa a "posição indecisa" na matriz completa original é irrelevante. Até porque poderíamos permutar a linha 3 por outra, e, então, y_1 ou y_2 passaria a ocupar tal posição. O y_3 não é "especial".

Exemplo 2.22
Sejam

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} -1 \\ 1 \\ 10 \end{bmatrix} \quad \text{e} \quad \mathbf{a}_4 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

Vamos, agora, investigar se o conjunto $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\}$ gera o $\mathbb{R}^3$.

Para isso, procedemos como no exemplo anterior, escalonando a matriz completa do sistema

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 + x_4\mathbf{a}_4 = \mathbf{y}, \tag{2.29}$$

onde $\mathbf{y}$ é, mais uma vez, um vetor qualquer de $\mathbb{R}^3$, de coordenadas y_1, y_2 e y_3.

$$\left[\begin{array}{cccc|c} \boxed{1} & 1 & -1 & 1 & y_1 \\ 2 & 3 & 1 & 0 & y_2 \\ -1 & 2 & 10 & 0 & y_3 \end{array}\right] \xrightarrow[\ell_3 \to \ell_3 + \ell_1]{\ell_2 \to \ell_2 - 2\ell_1} \left[\begin{array}{cccc|c} \boxed{1} & 1 & -1 & 1 & y_1 \\ 0 & \boxed{1} & 3 & -2 & y_2 - 2y_1 \\ 0 & 3 & 9 & 1 & y_3 + y_1 \end{array}\right] \longrightarrow$$

$$\xrightarrow{\ell_3 \to \ell_3 - 3\ell_2} \left[\begin{array}{cccc|c} \boxed{1} & 1 & -1 & 1 & y_1 \\ 0 & \boxed{1} & 3 & -2 & -2y_1 + y_2 \\ 0 & 0 & 0 & \boxed{7} & 7y_1 - 3y_2 + y_3 \end{array}\right] \tag{2.30}$$

A situação agora é bastante diferente daquela do exemplo anterior. A última coluna da matriz completa do sistema (2.29) *nunca* será pivô, não importa quais forem os valores de y_1, y_2 e y_3.[11] O sistema (2.29), portanto, é *sempre* possível, para qualquer vetor $\mathbf{y}$ de $\mathbb{R}^3$. Assim, o conjunto $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\}$ gera o $\mathbb{R}^3$.

Queremos caracterizar os conjuntos que geram $\mathbb{R}^n$, e usaremos os exemplos acima para nos guiarmos. Qual é a distinção essencial entre esses dois exemplos, no contexto desta questão?

Observe que a matriz de *coeficientes* $\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix}$ do sistema (2.27) possui uma linha sem posição-pivô: a terceira (ver (2.28)). Isso dá margem para que a última coluna da matriz *completa* $\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{y} \end{bmatrix}$ seja, para alguns valores de $\mathbf{y}$, uma coluna-pivô, como vimos no exemplo 2.21. Para tais valores de $\mathbf{y}$, o sistema (2.27) será impossível.

A matriz de coeficientes $\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 \end{bmatrix}$ do sistema (2.29), por outro lado, *possui uma posição-pivô em cada uma de suas três linhas* (ver (2.30)). Isso "fecha o cerco" de tal forma que a última coluna da matriz completa $\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 & \mathbf{y} \end{bmatrix}$ *nunca* poderá ser uma coluna-pivô. De fato, *cada linha* da forma escalonada em (2.30) já possui um elemento líder em uma coluna anterior à última, portanto, não há como "encaixar" mais um elemento líder nesta última coluna (verifique!). Dessa forma, a sistema (2.29) será sempre possível.

Isso motiva o resultado principal desta seção.

[11] Verifique esta afirmativa, examinando a forma escalonada em (2.30)!

Teorema 2.23
Sejam $\mathbf{a}_1$, $\mathbf{a}_2$, ... , $\mathbf{a}_m$ *vetores de* $\mathbb{R}^n$. *O conjunto* $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$ *gera o* $\mathbb{R}^n$ *se e somente se a matriz* $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_m \end{bmatrix}$ *possui uma posição-pivô em cada uma de suas* n linhas.

Demonstração: Por definição, o conjunto $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$ gera $\mathbb{R}^n$ se e somente se o sistema $A\mathbf{x} = \mathbf{y}$ é sempre possível, qualquer que seja o vetor $\mathbf{y}$ de $\mathbb{R}^n$ (lembre-se de que esse sistema é o mesmo que (2.26)). Vamos mostrar que o sistema $A\mathbf{x} = \mathbf{y}$ é sempre possível se e só se A possui uma posição-pivô em cada linha.

Seja F uma forma escalonada da matriz A. Dado um $\mathbf{y} \in \mathbb{R}^n$ qualquer, podemos escalonar a matriz completa do sistema $A\mathbf{x} = \mathbf{y}$:

$$\begin{bmatrix} A \mid \mathbf{y} \end{bmatrix} \xrightarrow[\text{(operações-linha)}]{\text{escalonamento}} \begin{bmatrix} F \mid \mathbf{z} \end{bmatrix}. \tag{2.31}$$

O vetor $\mathbf{z}$ é o vetor obtido aplicando-se sobre $\mathbf{y}$ as mesmas operações-linha que levam A até F. Observe que os sistemas $A\mathbf{x} = \mathbf{y}$ e $F\mathbf{x} = \mathbf{z}$ são equivalentes.

Se a matriz A possuir uma posição-pivô em cada linha, a matriz escalonada F terá um elemento líder em cada linha. A última coluna da matriz completa $\begin{bmatrix} F & \mathbf{z} \end{bmatrix}$, portanto, *nunca* será uma coluna-pivô, não importa qual seja o valor de $\mathbf{z}$ (ou de $\mathbf{y}$).[12] Sendo assim, o sistema $F\mathbf{x} = \mathbf{z}$ será *sempre possível*, e o mesmo valerá para o sistema equivalente $A\mathbf{x} = \mathbf{y}$.

Se, por outro lado, a matriz A *não* tiver uma posição-pivô em cada linha, então a *última linha* da matriz escalonada F será, necessariamente, toda de zeros.[13] Considere, então, um vetor $\mathbf{z}'$ de $\mathbb{R}^n$ cuja última coordenada z'_n seja diferente de zero ($z'_n = 1$, por exemplo). O sistema $F\mathbf{x} = \mathbf{z}'$ será, evidentemente, impossível, pois sua última equação será do tipo $0 = 1$. Agora considere o vetor $\mathbf{y}'$ de $\mathbb{R}^n$ obtido ao se aplicar sobre $\mathbf{z}'$ as operações-linha que *revertem* o processo (2.31) acima, isto é,

$$\begin{bmatrix} F \mid \mathbf{z}' \end{bmatrix} \xrightarrow[\text{(operações-linha reversas)}]{\text{“des-escalonamento”}} \begin{bmatrix} A \mid \mathbf{y}' \end{bmatrix}.$$

O sistema $A\mathbf{x} = \mathbf{y}'$ é equivalente ao sistema impossível $F\mathbf{x} = \mathbf{z}'$, portanto será, ele próprio, impossível. Assim, *não* é para qualquer escolha de $\mathbf{y}$ que o sistema $A\mathbf{x} = \mathbf{y}$ será possível. $\square$

Observe que, no caso em que a matriz A não tem uma posição-pivô em cada linha, a prova do teorema dá uma “receita” para a construção de um vetor $\mathbf{y}'$ de modo que o sistema $A\mathbf{x} = \mathbf{y}'$ seja impossível (essa dica poderá ajudá-los na resolução do exercício P2.31).

Mais uma vez, há muitas formas de se afirmar que um conjunto $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ gera $\mathbb{R}^n$. Para a conveniência do leitor, enumeramos várias delas a seguir.

Teorema 2.24
Sejam $\mathbf{a}_1$, $\mathbf{a}_2$, ... , $\mathbf{a}_m$ *vetores de* $\mathbb{R}^n$, *e seja* A *a matriz de tamanho* $n \times m$ *dada por* $\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_m \end{bmatrix}$. *As seguintes afirmativas são equivalentes:*

[12] A situação, nesse caso, é como a do exemplo 2.22: o “cerco está fechado”, e não há maneira de “encaixar” mais um elemento líder na última coluna de $\begin{bmatrix} F & \mathbf{z} \end{bmatrix}$.

[13] Do contrário, A teria uma posição-pivô em cada linha, não é mesmo? A situação aqui é como a do exemplo 2.21.

(a) $\operatorname{Span}\{\mathbf{a}_1, \dots, \mathbf{a}_m\} = \mathbb{R}^n$.

(b) *O conjunto* $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m\}$ *gera o* $\mathbb{R}^n$.

(c) *Os vetores* $\mathbf{a}_1$, $\mathbf{a}_2$, ... , $\mathbf{a}_m$ *geram o* $\mathbb{R}^n$.

(d) *Qualquer* $\mathbf{y} \in \mathbb{R}^n$ *é uma combinação linear dos vetores* $\mathbf{a}_1$, $\mathbf{a}_2$, ... , $\mathbf{a}_m$.

(e) *Para qualquer* $\mathbf{y} \in \mathbb{R}^n$, *existe ao menos uma lista de pesos* x_1, ... , x_m *tal que* $\mathbf{y} = x_1\mathbf{a}_1 + \cdots + x_m\mathbf{a}_m$.

(f) *Para qualquer* $\mathbf{y} \in \mathbb{R}^n$, *o sistema linear* $x_1\mathbf{a}_1 + \cdots + x_m\mathbf{a}_m = \mathbf{y}$ *é possível.*

(g) *Para qualquer* $\mathbf{y} \in \mathbb{R}^n$, *o sistema linear* $A\mathbf{x} = \mathbf{y}$ *é possível.*

(h) $\operatorname{Col} A = \mathbb{R}^n$.

(i) *A matriz* A *possui uma posição-pivô em cada uma de suas* n linhas.

A equivalência entre (b) e (i) é precisamente o teorema 2.23, e a equivalência entre as afirmativas (a) a (h) decorre, diretamente, das definições deste capítulo. Encorajamos que você faça a verificação de tais fatos. Essa tarefa será um bom exercício de fixação dos conceitos. As seguintes considerações podem ajudar.

A equivalência entre (a) e (d) é uma mera questão de linguagem: escrever $\operatorname{Span}\{\mathbf{a}_1, \dots, \mathbf{a}_m\} = \mathbb{R}^n$ se traduz, em palavras, para "o conjunto das combinações lineares (o *span*) de $\mathbf{a}_1$, ... , $\mathbf{a}_m$ é igual ao $\mathbb{R}^n$ todo", ou seja, qualquer vetor de $\mathbb{R}^n$ é uma combinação linear de $\mathbf{a}_1$, ... , $\mathbf{a}_m$. Sob a hipótese $A = \begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_m\end{bmatrix}$, vale $\operatorname{Col} A = \operatorname{Span}\{\mathbf{a}_1, \dots, \mathbf{a}_m\}$, portanto, a equivalência entre (a) e (h) é imediata.

Em geral, é a afirmativa (i) que é usada "na prática" para testar se todas as outras valem ou não. Ela é também a chave para demonstrar o seguinte resultado.

Proposição 2.25
Se os vetores $\mathbf{a}_1$, $\mathbf{a}_2$, ... , $\mathbf{a}_m$ *geram o* $\mathbb{R}^n$, *então, necessariamente, vale* $m \geqslant n$.

Demonstração: A matriz $n \times m$ dada por $A = \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_m\end{bmatrix}$ tem uma posição-pivô em cada uma de suas n linhas, uma vez que a afirmativa (i) do teorema 2.24 é equivalente à hipótese desta proposição. Isto implica que A tem exatamente n posições-pivô, pois uma matriz (qualquer que seja) não pode ter mais do que uma posição-pivô por linha.

O número m de colunas de A, portanto, *não pode ser menor do que* n. Do contrário, a matriz A não poderia ter n posições-pivô. Ela permitiria, no máximo, m posições-pivô, pois uma matriz qualquer também não pode ter mais do que uma posição-pivô por coluna. Veja o exercício P1.19 do capítulo 1. $\square$

A proposição 2.25 é simples, mas é importante. Uma outra forma de enunciá-la (a "contrapositiva") é dizer que se $m < n$, então os vetores $\mathbf{a}_1$, $\mathbf{a}_2$, ... , $\mathbf{a}_m$ não podem gerar $\mathbb{R}^n$ (e, portanto, não podem satisfazer nenhuma das afirmativas do teorema 2.24).

Isso é bastante intuitivo. Dois vetores de $\mathbb{R}^3$, por exemplo, nunca poderão gerar o $\mathbb{R}^3$ todo. Eles poderão gerar, no máximo, um *plano* dentro de $\mathbb{R}^3$, como no caso dos vetores $\mathbf{u}$ e $\mathbf{v}$ do exemplo 2.11 (ver também o exercício P2.22). Analogamente, dezesseis vetores, ou menos, em $\mathbb{R}^{17}$ jamais poderão gerar o $\mathbb{R}^{17}$ todo.

Cuidado!
As recíproca da proposição 2.25 *não* é verdadeira, ou seja, a condição $m \geqslant n$ *não garante* que valham as afirmativas do teorema 2.24. Em particular, três ou mais vetores de $\mathbb{R}^3$ não geram, necessariamente, o $\mathbb{R}^3$ (ver o exemplo 2.21). Similarmente, dezessete (ou mais) vetores em $\mathbb{R}^{17}$ não geram, necessariamente, o $\mathbb{R}^{17}$ todo. Para determinar se um dado conjunto de vetores gera $\mathbb{R}^n$, é necessário verificar diretamente uma das afirmativas do teorema 2.24 (geralmente a afirmativa (i), como mencionamos).

Exercícios propostos

P2.1. Sejam $\mathbf{x} = \left[\begin{smallmatrix} 3 \\ -2 \end{smallmatrix}\right]$, $\mathbf{y} = \left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$ e $\mathbf{z} = \left[\begin{smallmatrix} 0 \\ 4 \end{smallmatrix}\right]$. Calcule o vetor $2\mathbf{x} - 6\mathbf{y} + 3\mathbf{z}$.

P2.2. Represente um sistema de coordenadas cartesianas em uma folha de papel quadriculado. Posicione a origem bem no centro da folha, para que você tenha bastante espaço. Complete os itens abaixo, cuidadosamente, usando um par de esquadros.

(a) Represente os vetores $\mathbf{x}$, $\mathbf{y}$ e $\mathbf{z}$ do exercício anterior no sistema de coordenadas.
(b) Agora represente os vetores $2\mathbf{x}$, $-6\mathbf{y}$ e $3\mathbf{z}$.
(c) Obtenha, graficamente, o vetor $2\mathbf{x} + (-6\mathbf{y})$, usando a "regra do paralelogramo".
(d) Obtenha agora, graficamente, o vetor $\big(2\mathbf{x} + (-6\mathbf{y})\big) + 3\mathbf{z}$, usando, novamente, a regra do paralelogramo.
(e) Verifique que o vetor resultante, obtido graficamente, coincide com aquele obtido no exercício anterior.

P2.3. Calcule o produto $\begin{bmatrix} 2 & 5 & -3 & 1 \\ 0 & 1 & 4 & -2 \\ 3 & -4 & 0 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ -1 \\ 3 \end{bmatrix}$ de duas maneiras distintas:

(a) usando a definição 2.4, como no exemplo 2.5;
(b) usando o método "usual" empregado no ensino médio.

Perceba que os resultados são idênticos.

P2.4. Determine quais dos produtos abaixo estão definidos (alguns deles não fazem sentido). Calcule aqueles que estiverem.

(a) $\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \end{bmatrix}$

(b) $\begin{bmatrix} 3 & 1 & -6 \\ 2 & 0 & 3 \end{bmatrix} \begin{bmatrix} -1 \\ 3 \\ -2 \end{bmatrix}$

(c) $\begin{bmatrix} 4 & 1 \\ 2 & 0 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & -6 \\ 0 & 9 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \end{bmatrix}$

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

(f) $\begin{bmatrix} -3 & 5 & 7 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$

(g) $\begin{bmatrix} -3 & 5 & 7 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$

(h) $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

P2.5. Escreva a combinação linear $2\mathbf{x} - 6\mathbf{y} + 3\mathbf{z}$ do exercício P2.1 como o produto de uma matriz 2×3 por um vetor de $\mathbb{R}^3$. Observe que as coordenadas de tal vetor correspondem aos *pesos* na combinação linear.

P2.6. Seja A uma matriz $n \times m$ qualquer. Mostre que $A\mathbf{0}_m = \mathbf{0}_n$. (Lembre-se de que $\mathbf{0}_m$ é o vetor-zero de $\mathbb{R}^m$ e $\mathbf{0}_n$ é o vetor-zero de $\mathbb{R}^n$.) *Dica*: Use a definição do produto matriz-vetor, e verifique que $A\mathbf{0}_m$ é uma combinação linear com pesos todos iguais a zero.

P2.7. Seja $\mathbf{x}$ um vetor de $\mathbb{R}^m$ qualquer. Mostre que $\mathit{0}\mathbf{x} = \mathbf{0}_n$, onde $\mathit{0}$ é a matriz $n \times m$ com entradas todas iguais a zero.

P2.8. Verifique a propriedade (b) do produto de matriz por vetor (proposição 2.6, na página 44). *Sugestão*: Escreva A coluna por coluna e $\mathbf{u}$ coordenada por coordenada.

P2.9. Use o método de escalonamento para resolver o sistema linear (2.15) (página 45). Obtenha uma descrição vetorial paramétrica do conjunto-solução, como nos exemplos 2.7, 2.8 e 2.9.

P2.10. Considere os sistemas dos exercícios P1.10 e P1.11 do capítulo 1. Forneça uma descrição vetorial paramétrica do conjunto-solução de cada um que for *possível*. *Sugestão*: Aproveite o trabalho já realizado.

P2.11. Sejam $A = \begin{bmatrix} 4 & -1 & 3 \\ 1 & 2 & 0 \end{bmatrix}$ e $\mathbf{b} = \begin{bmatrix} 9 \\ 5 \end{bmatrix}$. Escreva a equação $A\mathbf{x} = \mathbf{b}$ explicitamente como um sistema de equações lineares. Esse sistema tem quantas variáveis? E quantas equações? Compare esses números com o tamanho da matriz A. Escreva a matriz completa do sistema explicitamente, e perceba que faz sentido denotá-la por $\begin{bmatrix} A \mid \mathbf{b} \end{bmatrix}$.

P2.12. Escreva o sistema (1.16) (página 19) na forma $A\mathbf{x} = \mathbf{b}$. Em outras palavras, determine A e $\mathbf{b}$ tais que o sistema (1.16) seja escrito nessa forma. Qual é o tamanho da matriz A? E do vetor $\mathbf{b}$? Compare com o número de variáveis e de equações do sistema (1.16).

P2.13. Em cada item a seguir, determine se o vetor $\mathbf{b}$ é uma combinação linear dos vetores $\mathbf{a}_1$, $\mathbf{a}_2$, ... , $\mathbf{a}_m$. Inspire-se nos exemplos 2.12 e 2.13.

(a) $\mathbf{a}_1 = \begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}$ e $\mathbf{b} = \begin{bmatrix} 0 \\ 14 \\ -2 \end{bmatrix}$.

(b) $\mathbf{a}_1 = \begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}$ e $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

(c) $\mathbf{a}_1 = \begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}$, $\mathbf{a}_3 = \begin{bmatrix} 0 \\ 14 \\ -2 \end{bmatrix}$ e $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

(d) $\mathbf{a}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} 2 \\ 3 \\ -1 \\ 0 \end{bmatrix}$, $\mathbf{a}_3 = \begin{bmatrix} 7 \\ 6 \\ 1 \\ 6 \end{bmatrix}$ e $\mathbf{b} = \begin{bmatrix} -3 \\ -6 \\ 3 \\ 2 \end{bmatrix}$.

P2.14. Um vetor de $\mathbb{R}^5$ pode ser uma combinação linear de vetores de $\mathbb{R}^8$? E um vetor de $\mathbb{R}^8$ pode pertencer ao *span* de vetores de $\mathbb{R}^5$? Justifique.

P2.15. Sejam $\mathbf{a}_1 = \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix}$ e $\mathbf{a}_2 = \begin{bmatrix} 1 \\ 1 \\ 7 \end{bmatrix}$. É verdade que $\operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$ é um subconjunto de $\mathbb{R}^2$? É um subconjunto de $\mathbb{R}^3$? Justifique.

P2.16. Seja $\mathbf{u} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$. O conjunto $\operatorname{Span}\{\mathbf{u}\}$ contém os vetores de $\mathbb{R}^2$ que podem ser escritos na forma $\alpha\mathbf{u}$, ou seja, $\operatorname{Span}\{\mathbf{u}\}$ é o conjunto de todos os múltiplos escalares de $\mathbf{u}$, certo?

(a) Verifique que um vetor $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ pertence a $\operatorname{Span}\{\mathbf{u}\}$ se e somente se suas coordenadas satisfazem a relação $3v_1 - 4v_2 = 0$. Lembre-se de que esta é a equação de uma reta no plano $\mathbb{R}^2$. Essa reta, portanto, é a representação geométrica de $\operatorname{Span}\{\mathbf{u}\}$.

(b) Represente $\mathbf{u}$, geometricamente, por uma seta em uma folha de papel quadriculado. Trace a reta $\operatorname{Span}\{\mathbf{u}\}$ no mesmo esboço. Observe que ela é a reta que "contém a seta" $\mathbf{u}$.

P2.17. Sejam $\mathbf{u} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$ e $\mathbf{v} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$.

(a) Verifique que $\mathbf{u}$ e $\mathbf{v}$ geram $\mathbb{R}^2$. *Dica*: Use o teorema 2.23.

(b) Relembre que o ítem anterior equivale a dizer $\operatorname{Span}\{\mathbf{u}, \mathbf{v}\} = \mathbb{R}^2$ (teorema 2.24).

(c) Conclua que $\operatorname{Span}\{\mathbf{u}, \mathbf{v}\}$ é representado geometricamente pelo plano $\mathbb{R}^2$ todo.

(d) Represente os vetores $\mathbf{u}$ e $\mathbf{v}$ em uma folha de papel quadriculado, e observe que esses vetores não são colineares.

P2.18. Verifique que o subespaço gerado pelo vetor zero de $\mathbb{R}^n$ é o subconjunto de $\mathbb{R}^n$ que contém apenas o próprio vetor zero, isto é, $\operatorname{Span}\{\mathbf{0}\} = \{\mathbf{0}\}$. *Dica*: Quais são os múltiplos do vetor zero?

P2.19. Dizemos que dois vetores $\mathbf{u}$ e $\mathbf{v}$ de $\mathbb{R}^n$ são *colineares* quando $\mathbf{u}$ é um múltiplo escalar de $\mathbf{v}$ *ou vice-versa*. A figura 2.6(b) ilustra um exemplo em $\mathbb{R}^3$: os vetores $\mathbf{u}'$ e $\mathbf{v}'$ são colineares.

(a) Suponha que $\mathbf{u}$ e $\mathbf{v}$ sejam *não-nulos*. Mostre que $\mathbf{u}$ é um múltiplo escalar de $\mathbf{v}$ se e somente se $\mathbf{v}$ é um múltiplo escalar de $\mathbf{u}$. *Dica*: Escreva $\mathbf{u} = k\mathbf{v}$ para algum escalar k, e argumente, usando as hipóteses, que $k \neq 0$.

(b) Agora, suponha que $\mathbf{u}$ seja não-nulo, mas que $\mathbf{z} = \mathbf{0}$. Mostre que esses vetores são "automaticamente" colineares. (*Dica*: Verifique que $\mathbf{z}$ é um múltiplo de $\mathbf{u}$.) Mostre, no entanto, que $\mathbf{u}$ *não é um múltiplo de* $\mathbf{z}$.

P2.20. (a) Suponha que $\mathbf{u}'$ e $\mathbf{v}'$ sejam vetores *colineares* e *não-nulos* em $\mathbb{R}^n$ (como na figura 2.6(b)). Mostre que qualquer combinação linear $x_1\mathbf{u}' + x_2\mathbf{v}'$ pode ser escrita na forma $\alpha\mathbf{u}'$ e na forma $\beta\mathbf{v}'$. Em outras palavras, qualquer combinação linear de $\mathbf{u}'$ e $\mathbf{v}'$ é, simplesmente, um múltiplo escalar de $\mathbf{u}'$ (ou de $\mathbf{v}'$). Isso equivale a dizer $\operatorname{Span}\{\mathbf{u}', \mathbf{v}'\} = \operatorname{Span}\{\mathbf{u}'\} = \operatorname{Span}\{\mathbf{v}'\}$, certo?

(b) Agora, suponha que $\mathbf{u}'$ seja não-nulo, mas que $\mathbf{v}' = \mathbf{0}$. Mostre que $\operatorname{Span}\{\mathbf{u}', \mathbf{v}'\} = \operatorname{Span}\{\mathbf{u}'\}$, mas que $\operatorname{Span}\{\mathbf{v}'\}$ *não* coincide com esse conjunto. De fato, neste caso $\operatorname{Span}\{\mathbf{v}'\} = \{\mathbf{0}\}$, conforme o exercício P2.18.

P2.21. Considere os vetores $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{a}_3$ do exemplo 2.21 (página 54). Neste exercício, vamos descrever geometricamente o conjunto $\operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$.

(a) Mostre que um vetor $\mathbf{y} \in \mathbb{R}^3$ é gerado por $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{a}_3$ se e somente se suas coordenadas satisfazem a relação

$$7y_1 - 3y_2 + y_3 = 0. \tag{2.32}$$

Dica: Aproveite o trabalho já desenvolvido no exemplo 2.21.

(b) Recorde-se (do ensino médio, ou de um curso de geometria analítica) de que (2.32) é a equação de um *plano* em $\mathbb{R}^3$ passando pela origem.

(c) Convença-se (usando as definições deste capítulo) que o item (a) deste exercício pode ser reformulado da seguinte maneira: O *span* de $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{a}_3$ é o conjunto dos vetores de $\mathbb{R}^3$ que estão no plano dado por (2.32). Ou, mais sucintamente: $\operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ é o plano dado por (2.32).

(d) Verifique que cada um dos vetores $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{a}_3$ pertence ao plano (2.32) (e não poderia ser de outra maneira, em vista da observação 2.15). Esses vetores, portanto, são *coplanares*: os três estão contidos num mesmo plano. Observe, no entanto, que eles *não são* colineares.

A figura 2.7 ilustra essa situação, em que os três vetores $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{a}_3$ de $\mathbb{R}^3$ geram um plano passando pela origem.[14] Denotamos o plano por W.

[14] A caixa pontilhada está na figura para realçar sua "tridimensionalidade", mas note que os eixos x_1 e x_2 não atravessam as faces laterais perpendicularmente (isto é, a caixa é "torta" com relação aos eixos).

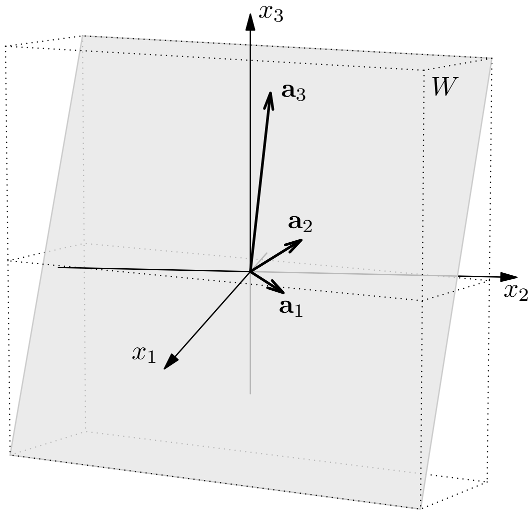


Figura 2.7: O plano W gerado por $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{a}_3$ (veja o exercício P2.21).

P2.22. Convença-se dos fatos abaixo. Demonstrações rigorosas não são necessárias aqui, pois o objetivo é apenas o de desenvolver a intuição geométrica. Faça esboços, imagens mentais, ou gráficos auxiliados por computador.

(a) Se $\mathbf{u} \in \mathbb{R}^2$ é um vetor não-nulo, então $\operatorname{Span}\{\mathbf{u}\}$ é uma reta contendo a origem no plano $\mathbb{R}^2$ (ver o exercício P2.16).
(b) Se $\mathbf{u}$ e $\mathbf{v}$ são vetores não-colineares em $\mathbb{R}^2$, então $\operatorname{Span}\{\mathbf{u}, \mathbf{v}\}$ é o plano $\mathbb{R}^2$ todo (ver o exercício P2.17).
(c) Se $\mathbf{u}$ e $\mathbf{v}$ são vetores colineares em $\mathbb{R}^2$, então $\operatorname{Span}\{\mathbf{u}, \mathbf{v}\}$ é uma reta contendo a origem no plano $\mathbb{R}^2$.
(d) Se $\mathbf{u} \in \mathbb{R}^3$ é um vetor não-nulo, então $\operatorname{Span}\{\mathbf{u}\}$ é uma reta contendo a origem no espaço $\mathbb{R}^3$.
(e) Se $\mathbf{u}$ e $\mathbf{v}$ são vetores não-colineares em $\mathbb{R}^3$, então $\operatorname{Span}\{\mathbf{u}, \mathbf{v}\}$ é um plano contendo a origem no espaço $\mathbb{R}^3$ (ver a figura 2.6(a)).
(f) Se $\mathbf{u}'$ e $\mathbf{v}'$ são vetores colineares em $\mathbb{R}^3$, então $\operatorname{Span}\{\mathbf{u}', \mathbf{v}'\}$ é uma reta contendo a origem no espaço $\mathbb{R}^3$ (ver a figura 2.6(b)).
(g) Se $\mathbf{u}$, $\mathbf{v}$ e $\mathbf{w}$ são vetores não-coplanares em $\mathbb{R}^3$, então $\operatorname{Span}\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ é o espaço $\mathbb{R}^3$ todo.
(h) Se $\mathbf{u}$, $\mathbf{v}$ e $\mathbf{w}$ são vetores coplanares, porém não-colineares em $\mathbb{R}^3$, então $\operatorname{Span}\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ é um plano contendo a origem no espaço $\mathbb{R}^3$ (ver o exercício P2.21).
(i) Se $\mathbf{u}$, $\mathbf{v}$ e $\mathbf{w}$ são vetores colineares em $\mathbb{R}^3$, então $\operatorname{Span}\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ é uma reta contendo a origem no espaço $\mathbb{R}^3$.

P2.23. Seja $A = \begin{bmatrix} 3 & 2 \\ 2 & -1 \\ 7 & 4 \end{bmatrix}$. Em cada item, determine se o vetor dado pertence a $\operatorname{Col} A$. Inspire-se no exemplo 2.18 e na proposição 2.19.

(a) $\mathbf{u} = \begin{bmatrix} -3 \\ 5 \\ -5 \end{bmatrix}$. (b) $\mathbf{v} = \begin{bmatrix} 2 \\ 2 \\ 5 \end{bmatrix}$.

P2.24. É possível resolver, simultaneamente, ambos os itens da questão anterior, via o escalonamento da matriz $\begin{bmatrix} A & \mathbf{u} & \mathbf{v} \end{bmatrix}$. Explique como isso é possível. *Sugestão*: Primeiro, faça o escalonamento proposto.

P2.25. Seja B uma matriz 6×4. $\operatorname{Col} B$ é um subconjunto de qual "espaço ambiente"? De $\mathbb{R}^6$ ou de $\mathbb{R}^4$?

P2.26. Nas afirmativas abaixo, A representa uma matriz $n \times m$. Determine se cada uma é verdadeira ou falsa. Justifique.

(a) *Sempre* vale $\operatorname{Col} A = \mathbb{R}^n$.
(b) $\operatorname{Col} A$ é *sempre* um subconjunto de $\mathbb{R}^n$.
(c) $\operatorname{Col} A$ é *sempre* um subconjunto de $\mathbb{R}^m$.
(d) Cada coluna de A pertence ao seu espaço-coluna $\operatorname{Col} A$. *Dica*: Veja a observação 2.15.
(e) O espaço-coluna de A é o conjunto que contém *apenas* as colunas de A.
(f) O espaço-coluna de A é o conjunto gerado por seus vetores-coluna.
(g) Se $\mathbf{y} \in \operatorname{Col} A$, então $\mathbf{y}$ é um vetor de $\mathbb{R}^n$ que pode ser escrito como uma combinação linear dos vetores-coluna de A.

P2.27. Seja C a matriz $\begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_p \end{bmatrix}$, onde $\mathbf{c}_j$ são vetores de $\mathbb{R}^q$, e seja $\mathbf{z}$ um vetor qualquer de $\mathbb{R}^q$. Qual é o tamanho da matriz C? Escreva uma lista com diversas maneiras distintas de dizer "$\mathbf{z} \in \operatorname{Col} C$".

P2.28. Sejam $\mathbf{a}_1 = \begin{bmatrix} 3 \\ 2 \\ 7 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix}$, $\mathbf{a}_3 = \begin{bmatrix} -3 \\ 5 \\ -5 \end{bmatrix}$ e $\mathbf{a}_4 = \begin{bmatrix} 2 \\ 2 \\ 5 \end{bmatrix}$. Determine quais dos conjuntos abaixo geram $\mathbb{R}^3$. *Dica*: Tente reaproveitar o trabalho já realizado no exercício P2.23.

(a) $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$, (b) $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_4\}$, (c) $\{\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_4\}$.

P2.29. Sejam $A = \begin{bmatrix} 5 & 4 & -3 \\ 0 & -3 & 6 \\ 2 & 1 & 0 \end{bmatrix}$ e $B = \begin{bmatrix} 5 & 4 & -3 & 1 \\ 0 & -3 & 6 & 1 \\ 2 & 1 & 0 & 1 \end{bmatrix}$.

(a) Determine se vale $\operatorname{Col} A = \mathbb{R}^3$. O que isso diz a respeito dos vetores-coluna de A?
(b) Determine se vale $\operatorname{Col} B = \mathbb{R}^3$. Vale $\operatorname{Col} B = \mathbb{R}^4$?

P2.30. Sejam $\mathbf{x} = \left[\begin{smallmatrix} 1 \\ 3 \\ 2 \end{smallmatrix}\right]$ e $\mathbf{y} = \left[\begin{smallmatrix} 2 \\ 4 \\ 1 \end{smallmatrix}\right]$. Nos itens abaixo, justifique suas respostas *sem fazer conta alguma*.

(a) O conjunto $\{\mathbf{x}, \mathbf{y}\}$ gera o $\mathbb{R}^3$?
(b) O conjunto $\{\mathbf{x}, \mathbf{y}\}$ gera o $\mathbb{R}^2$?

P2.31. Considere a matriz $A = \begin{bmatrix} 2 & 6 & 1 \\ 3 & 9 & 1 \\ -1 & -3 & 4 \end{bmatrix}$.

(a) Verifique que as colunas de A *não* geram o $\mathbb{R}^3$.
(b) Dê exemplos de vetores de $\mathbb{R}^3$ que não pertençam a $\operatorname{Col} A$, isto é, vetores que não sejam gerados pelas colunas de A.

Dica: Inspire-se no exemplo 2.21 (ou na prova do teorema 2.23).

P2.32. As colunas de uma matriz 100×99 podem gerar o conjunto $\mathbb{R}^{100}$? Justifique.

P2.33. Determine se cada afirmativa é verdadeira ou falsa. Justifique.

(a) Quatro vetores de $\mathbb{R}^5$ *nunca* geram o $\mathbb{R}^5$.
(b) Cinco vetores de $\mathbb{R}^5$ *sempre* geram o $\mathbb{R}^5$.
(c) Cinco vetores de $\mathbb{R}^5$ *nunca* geram o $\mathbb{R}^5$.
(d) Um conjunto contendo *exatamente cinco* vetores pode gerar o $\mathbb{R}^5$.
(e) Um conjunto com *mais* de cinco vetores pode gerar o $\mathbb{R}^5$.
(f) Quatro vetores de $\mathbb{R}^5$ nunca geram o $\mathbb{R}^5$, mas podem gerar o $\mathbb{R}^4$.

Capítulo 3

Sistemas Homogêneos, o Núcleo de uma Matriz e Independência Linear

3.0 Introdução

Neste capítulo, introduziremos três conceitos importantes, que estão intimamente relacionados.

Na primeira seção, estudaremos sistemas lineares cujos termos independentes são todos iguais a zero. Tais sistemas, ditos *homogêneos*, têm certo papel especial em álgebra linear. Na segunda seção, estudaremos mais a fundo os conjuntos-solução desses sistemas, abordando, também, seus aspectos geométricos.

Finalmente, abordaremos o conceito fundamental de *independência linear*, na última seção. Essencialmente, um conjunto de vetores é linearmente independente quando nenhum de seus elementos pode ser escrito como uma combinação linear dos demais. O conceito será "oficialmente" definido de uma outra maneira, mais conveniente. Veremos, no entanto, que a definição formal será equivalente a essa caracterização intuitiva (conforme a proposição 3.18).

3.1 Sistemas lineares homogêneos

Lembre-se de que qualquer sistema de n equações lineares com m variáveis pode ser escrito na "forma compacta" $A\mathbf{x} = \mathbf{b}$, onde A é uma matriz $n \times m$ (a matriz de coeficientes do sistema), e $\mathbf{b}$ é um vetor de $\mathbb{R}^n$ (o vetor dos termos independentes). A definição abaixo irá distinguir o caso especial em que os termos independentes são todos iguais a zero, isto é, o caso em que $\mathbf{b}$ é o vetor zero $\mathbf{0}_n$.

Definição 3.1
Um sistema linear é dito **homogêneo** se puder ser escrito na forma $A\mathbf{x} = \mathbf{0}_n$, onde A é uma matriz $n \times m$.

O sistema $A\mathbf{x} = \mathbf{b}$ é dito **não-homogêneo** se $\mathbf{b}$ *não* for o vetor zero.

Os sistemas (a), (b) e (c) da página 1 são todos não-homogêneos, bem como quase todos os outros exemplos que vimos até aqui. Já os sistemas

$$\begin{cases} x_1 - 3x_2 + 2x_3 = 0 \\ 2x_1 + x_2 - x_3 = 0 \end{cases} \quad \text{e} \quad \begin{cases} x_1 - 1 + 2x_3 = 3x_2 - 1 \\ x_2 + 2 - x_3 = 2 - 2x_1 \end{cases} \tag{3.1}$$

são homogêneos. O segundo “parece” não-homogêneo, mas isso é enganoso. É fácil ver que esses sistemas são equivalentes e podem ser escritos na forma

$$\begin{bmatrix} 1 & -3 & 2 \\ 2 & 1 & -1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tag{3.2}$$

onde $\mathbf{x} \in \mathbb{R}^3$ é o vetor das variáveis x_1, x_2 e x_3 (exercício P3.1). A matriz completa desse sistema é

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

A matriz completa do sistema homogêneo “geral” $A\mathbf{x} = \mathbf{0}_n$ é dada por $\left[A \mid \mathbf{0}_n\right]$. Perceba que esta notação faz sentido.

Sistemas homogêneos são *sempre* possíveis. De fato, qualquer que seja a matriz A ($n \times m$), o vetor zero $\mathbf{0}_m$ de $\mathbb{R}^m$ será *sempre* uma solução do sistema $A\mathbf{x} = \mathbf{0}_n$, pois $A\mathbf{0}_m = \mathbf{0}_n$ (ver o exercício P2.6). Chamamos $\mathbf{x} = \mathbf{0}_m$ de **solução trivial** do sistema $A\mathbf{x} = \mathbf{0}_n$. A seguir iremos sintetizar essas observações.

Observação 3.2
Seja A uma matriz $n \times m$ qualquer. O sistema homogêneo $A\mathbf{x} = \mathbf{0}_n$ *sempre* possui a solução trivial $\mathbf{x} = \mathbf{0}_m$. Em particular, todo sistema homogêneo é possível.

Um sistema homogêneo $A\mathbf{x} = \mathbf{0}$ pode ou não ter soluções *não-triviais*, isto é, soluções diferentes do vetor zero.[1] A proposição a seguir caracteriza esta questão em termos das posições-pivô da matriz A, e será fundamental na seção 3.3.

Proposição 3.3
Um sistema linear homogêneo $A\mathbf{x} = \mathbf{0}$ *ou possui unicamente a solução trivial, ou então possui uma infinidade de soluções (a trivial e mais uma infinidade de soluções não-triviais).*

O sistema $A\mathbf{x} = \mathbf{0}$ *possui* unicamente *a solução trivial* $\mathbf{x} = \mathbf{0}$ *se e somente se todas as colunas de sua matriz de coeficientes A são colunas-pivô.*

Demonstração: Esta é uma consequência direta do teorema de existência e unicidade 1.13. Já sabemos que o sistema $A\mathbf{x} = \mathbf{0}$ é possível, isto é, possui ao menos uma solução. Se todas as colunas da matriz A forem colunas-pivô, então o sistema $A\mathbf{x} = \mathbf{0}$ não terá variáveis livres, e, portanto, irá possuir uma única solução. Como $\mathbf{x} = \mathbf{0}$ é uma solução, ela será, exata e necessariamente, esta única solução!

Se, por outro lado, a matriz A tiver pelo menos uma coluna não-pivô, então o sistema $A\mathbf{x} = \mathbf{0}$ terá pelo menos uma variável livre, e, portanto, irá possuir uma infinidade de soluções. Sendo assim, o sistema $A\mathbf{x} = \mathbf{0}$ necessariamente terá uma infinidade de soluções, *além da solução-trivial* $\mathbf{x} = \mathbf{0}$. □

[1] Para simplificar a notação, iremos usar “$\mathbf{0}$” para indicar os vetores $\mathbf{0}_n$ e $\mathbf{0}_m$. O contexto deverá ser suficiente para que você faça a distinção.

Exemplo 3.4
Vamos determinar se o sistema homogêneo $B\mathbf{x} = \mathbf{0}$ possui soluções não-triviais, onde

$$B = \begin{bmatrix} 1 & 0 & 1 \\ 3 & 4 & -5 \\ 0 & -1 & 2 \end{bmatrix}. \tag{3.3}$$

Escalonemos a matriz de coeficientes B a fim de localizar suas posições-pivô:

$$B = \begin{bmatrix} \boxed{1} & 0 & 1 \\ 3 & 4 & -5 \\ 0 & -1 & 2 \end{bmatrix} \xrightarrow{\ell_2 \to \ell_2 - 3\ell_1} \begin{bmatrix} \boxed{1} & 0 & 1 \\ 0 & \boxed{4} & -8 \\ 0 & -1 & 2 \end{bmatrix} \xrightarrow{\ell_3 \to \ell_3 + \frac{1}{4}\ell_2} \begin{bmatrix} \boxed{1} & 0 & 1 \\ 0 & \boxed{4} & -8 \\ 0 & 0 & 0 \end{bmatrix}.$$

Não é necessário obter a forma escalonada reduzida. Observe que B possui uma coluna não-pivô (a terceira). Pela proposição 3.3, o sistema $B\mathbf{x} = \mathbf{0}$ possui uma infinidade de soluções não-triviais, além da solução trivial.

Isso responde à questão proposta nesse exemplo, mas desejamos explorá-lo um pouco mais, a fim de elucidar a proposição 3.3. Vamos, então, *resolver* o sistema $B\mathbf{x} = \mathbf{0}$, e obter explicitamente todas as suas soluções. Verifique que a forma escalonada reduzida de sua matriz completa $\begin{bmatrix} B \mid \mathbf{0} \end{bmatrix}$ é

$$\left[\begin{array}{ccc|c} \boxed{1} & 0 & 1 & 0 \\ 0 & \boxed{1} & -2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right], \tag{3.4}$$

Observe que x_3 é uma variável livre do sistema $B\mathbf{x} = \mathbf{0}$, e que uma descrição vetorial paramétrica de seu conjunto-solução (ver seção 2.4) é dada por

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -x_3 \\ 2x_3 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}. \tag{3.5}$$

Agora é evidente que o sistema possui uma infinidade de soluções: todos os múltiplos do vetor $\left[\begin{smallmatrix} -1 \\ 2 \\ 1 \end{smallmatrix}\right]$. Para cada escolha de $x_3 \neq 0$ em (3.5), obtemos uma solução não-trivial distinta. Fazendo $x_3 = 0$, obtemos a solução trivial $\mathbf{x} = \mathbf{0}$.

Exemplo 3.5
Agora vamos determinar se o sistema homogêneo $C\mathbf{x} = \mathbf{0}$ possui soluções não-triviais, onde

$$C = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

Basta uma operação-linha para escalonar a matriz C:

$$C = \begin{bmatrix} \boxed{1} & 2 \\ 3 & 4 \end{bmatrix} \xrightarrow{\ell_2 \to \ell_2 - 3\ell_1} \begin{bmatrix} \boxed{1} & 2 \\ 0 & \boxed{-2} \end{bmatrix}.$$

Pela proposição 3.3, o sistema $C\mathbf{x} = \mathbf{0}$ possui somente a solução trivial $\mathbf{x} = \mathbf{0}$, já que ambas as colunas de C são colunas-pivô. Como exercício, sugerimos que você resolva o sistema $C\mathbf{x} = \mathbf{0}$ usando o método de escalonamento, e, assim, obtenha uma verificação direta de que $\mathbf{x} = \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]$ é a única solução. Ao fazer esse exercício, repare, em particular, que o sistema não tem variáveis livres.

Um sistema linear homogêneo com mais variáveis do que equações tem, necessariamente, variáveis livres, e, portanto, uma infinidade de soluções. Enunciamos esse resultado, formalmente, a seguir. Lembre que se A é uma matriz $n \times m$, então m é o número de variáveis do sistema $A\mathbf{x} = \mathbf{0}$ e n o de equações.

Proposição 3.6
Seja A uma matriz $n \times m$. Se $m > n$, então o sistema homogêneo $A\mathbf{x} = \mathbf{0}$ possui uma infinidade de soluções (em particular, de soluções não-triviais).

Demonstração: Como A tem n linhas, essa matriz tem *no máximo* n posições-pivô. Já que A tem m colunas, e $m > n$, então, necessariamente, há colunas sem posição-pivô. Assim sendo, pela proposição 3.3, o sistema $A\mathbf{x} = \mathbf{0}$ tem uma infinidade de soluções. $\square$

Exemplo 3.7
Consideremos o seguinte "sistema" homogêneo de uma só equação:

$$2x_1 + 5x_2 - 6x_3 = 0. \tag{3.6}$$

Usando a notação vetorial, essa equação se escreve na forma $D\mathbf{x} = 0$, onde $D = \begin{bmatrix} 2 & 5 & -6 \end{bmatrix}$ (verifique).[2] Pela proposição 3.6, o "sistema" (3.6) possui uma infinidade de soluções, já que D é uma matriz 1×3 e $3 > 1$ (ou, em termos mais simples, já que (3.6) é um sistema homogêneo com mais variáveis do que equações). Neste exemplo, é fácil verificar isso diretamente, pois as soluções de (3.6) são dadas por

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -\frac{5}{2}x_2 + 3x_3 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} -\frac{5}{2} \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}. \tag{3.7}$$

Obtemos uma solução distinta para cada escolha das variáveis livres x_2 e x_3.

3.2 O núcleo de uma matriz

O conjunto-solução de um sistema homogêneo é um "objeto matemático" importante, pois aparece naturalmente em muitos problemas de álgebra linear e aplicações. Veremos exemplos disso em alguns capítulos mais adiante. Em virtude de sua importância, há um nome especial para esse "objeto", conforme a definição a seguir.

Definição 3.8
Seja A uma matriz de tamanho $n \times m$. O conjunto-solução do sistema linear homogêneo $A\mathbf{x} = \mathbf{0}_n$ é um subconjunto de $\mathbb{R}^m$ que chamamos de **núcleo** ou **espaço nulo** da matriz A. Denotamos o núcleo de A por $\operatorname{Nuc} A$.

Ou seja, um vetor $\mathbf{u}$ de $\mathbb{R}^m$ pertence ao núcleo de A se e somente se $A\mathbf{u} = \mathbf{0}_n$.

[2] Repare que escrevemos o zero *escalar* em $D\mathbf{x} = 0$. O vetor zero $\mathbf{0}$ de $\mathbb{R}^1$ corresponde simplesmente ao escalar 0, quando pensamos no conjunto $\mathbb{R}^1$ como o conjunto $\mathbb{R}$ dos escalares.

Alguns autores denotam o núcleo de A como $\operatorname{Nul} A$ ou $\operatorname{Ker} A$, ao invés de $\operatorname{Nuc} A$. A notação $\operatorname{Ker} A$ é padrão em textos em inglês e alemão, pois, nessas línguas, as palavras usadas para *núcleo* são *kernel* e *Kern*, respectivamente.

Cuidado!
Não confunda o núcleo de uma matriz com seu espaço-coluna. Ambos são subconjuntos definidos em termos de uma matriz dada, mas veja que suas definições são bastante diferentes! Observe, em particular, que se a matriz A é $n \times m$, então $\operatorname{Col} A$ é um subconjunto de $\mathbb{R}^n$ e $\operatorname{Nuc} A$ é um subconjunto de $\mathbb{R}^m$.

Exemplo 3.9
Seja $A = \begin{bmatrix} 1 & -3 & 2 \\ 2 & 1 & -1 \end{bmatrix}$. Determine se os vetores $\mathbf{u} = \left[\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right]$, $\mathbf{v} = \left[\begin{smallmatrix} 1 \\ 5 \\ 7 \end{smallmatrix}\right]$ e $\mathbf{0}_3 = \left[\begin{smallmatrix} 0 \\ 0 \\ 0 \end{smallmatrix}\right]$ pertencem ao núcleo de A.

Solução: Verifique que $A\mathbf{u} = \left[\begin{smallmatrix} 0 \\ 2 \end{smallmatrix}\right]$. Como este não é o vetor zero, $\mathbf{u}$ *não* pertence a $\operatorname{Nuc} A$. Por outro lado, vale $A\mathbf{v} = \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right] = \mathbf{0}_2$, logo $\mathbf{v} \in \operatorname{Nuc} A$. Finalmente, é fácil ver que $A\mathbf{0}_3 = \mathbf{0}_2$, logo o vetor zero $\mathbf{0}_3$ também pertence ao núcleo de A. □

Observação 3.10
Seja A uma matriz $n \times m$ qualquer. O núcleo de A *sempre* contém o vetor zero $\mathbf{0}_m$ de $\mathbb{R}^m$. Em particular, $\operatorname{Nuc} A$ *nunca* é o conjunto vazio.

Essa observação é uma mera reformulação, em termos do conceito de núcleo, da observação 3.2. Podemos reformular também a importante proposição 3.3.

Proposição 3.11
Seja A uma matriz qualquer. O núcleo de A contém unicamente o vetor zero (em símbolos, $\operatorname{Nuc} A = \{\mathbf{0}\}$*) se e somente se todas as colunas de A são colunas-pivô.*

O exemplo 3.9 mostra que verificar se um dado vetor pertence ao núcleo de uma matriz é uma tarefa muito simples. O núcleo de A é descrito, *implicitamente*, pela relação $A\mathbf{x} = \mathbf{0}$.[3] Assim, para determinar se um dado vetor pertence a $\operatorname{Nuc} A$, basta "testar" se ele satisfaz ou não essa equação.

Dada uma matriz A, no entanto, como podemos encontrar uma descrição *explícita* de seu núcleo? Ora, isso também é simples. Basta *resolver* o sistema $A\mathbf{x} = \mathbf{0}$, e obter uma descrição paramétrica de seu conjunto-solução, que, por definição, é $\operatorname{Nuc} A$.

Exemplo 3.12
Obtenha uma descrição explícita do núcleo da matriz B do exemplo 3.4.

Solução: Teríamos que resolver o sistema $B\mathbf{x} = \mathbf{0}$, mas já fizemos isso no exemplo 3.4. Obtivemos, em (3.5), uma descrição explícita de $\operatorname{Nuc} B$, em forma vetorial paramétrica.

Queremos apresentar esse resultado de uma maneira mais formal. A descrição (3.5) diz que $\operatorname{Nuc} B$ é o conjunto dos múltiplos do vetor $\mathbf{u} = \left[\begin{smallmatrix} -1 \\ 2 \\ 1 \end{smallmatrix}\right]$. Isso

[3] Dizemos isso porque a condição $A\mathbf{x} = \mathbf{0}$ não fornece uma "lista" ou descrição *explícita* dos elementos de $\operatorname{Nuc} A$.

se traduz, em símbolos, para $\operatorname{Nuc} B = \operatorname{Span}\{\mathbf{u}\}$ (faça uma rápida revisão da seção 2.5, se necessário).

$\operatorname{Span}\{\mathbf{u}\}$ é a descrição explícita do núcleo de B que procurávamos. Vale a pena, no entanto, fazer mais alguns comentários de natureza geométrica acerca deste exemplo. O conjunto $\operatorname{Nuc} B = \operatorname{Span}\{\mathbf{u}\}$ é representado pela *reta* no $\mathbb{R}^3$ que passa pela origem e que contém o vetor $\mathbf{u}$.

Reveja a matriz B em (3.3) e verifique que $B\mathbf{x} = \mathbf{0}$ equivale ao sistema

$$\begin{cases} x_1 \phantom{{}+4x_2} + x_3 = 0 \\ 3x_1 + 4x_2 - 5x_3 = 0 \\ - x_2 + 2x_3 = 0. \end{cases} \tag{3.8}$$

Cada uma das três equações acima representa (implicitamente) um *plano passando pela origem* em $\mathbb{R}^3$. A interseção desses três planos é justamente o conjunto-solução do sistema (3.8), ou seja, é $\operatorname{Nuc} B$. Como acabamos de discutir, essa interseção $\operatorname{Nuc} B$ é uma reta. Tente visualizar esta situação: três planos no espaço cuja interseção seja uma reta.

Enfatizamos que $\operatorname{Span}\{\mathbf{u}\}$ é uma descrição *explícita* da reta $\operatorname{Nuc} B$, enquanto (3.8) (ou, equivalentemente, $B\mathbf{x} = \mathbf{0}$) é uma descrição *implícita* da mesma. □

Exemplo 3.13
Obtenha uma descrição explícita do núcleo da matriz $D = \begin{bmatrix} 2 & 5 & -6 \end{bmatrix}$.

Solução: Já fizemos isso no exemplo 3.7. Obtivemos, em (3.7), uma descrição explícita do conjunto-solução de $D\mathbf{x} = 0$, isto é, de $\operatorname{Nuc} D$.

Novamente, desejamos escrever o resultado de maneira mais formal. A descrição (3.7) mostra que $\operatorname{Nuc} D$ é o conjunto das combinações lineares dos vetores $\mathbf{v}_1 = \begin{bmatrix} -5/2 \\ 1 \\ 0 \end{bmatrix}$ e $\mathbf{v}_2 = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}$. Em símbolos, temos $\operatorname{Nuc} D = \operatorname{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$.

Geometricamente, $\operatorname{Nuc} D$ é representado pelo *plano* do $\mathbb{R}^3$ gerado pelos vetores $\mathbf{v}_1$ e $\mathbf{v}_2$. Observamos, como no exemplo anterior, que $2x_1 + 5x_2 - 6x_3 = 0$ (ou, equivalentemente, $D\mathbf{x} = 0$) é uma descrição implícita do plano $\operatorname{Nuc} D$. O mesmo plano é descrito explicitamente por $\operatorname{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$. □

Dada uma matriz A, é sempre possível escrever $\operatorname{Nuc} A$, explicitamente, como o *span* de certos vetores[4], como fizemos nos exemplos acima. *Esse procedimento será muito usado em outras seções, problemas e aplicações, e, por isso, recomendamos fortemente o exercício* P3.10.

É fácil interpretar, geometricamente, o núcleo de uma matriz quando este é um subconjunto de $\mathbb{R}^2$ ou $\mathbb{R}^3$, pois, nesses casos, uma concepção visual direta é possível.

3.3 Dependência e independência linear

Considere os vetores $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ e $\mathbf{v} = \begin{bmatrix} 6 \\ 3 \end{bmatrix}$ de $\mathbb{R}^2$. Note que $\mathbf{u}$ e $\mathbf{v}$ são colineares: $\mathbf{v}$ é um múltiplo de $\mathbf{u}$, a saber, $\mathbf{v} = 3\mathbf{u}$. O vetor $\mathbf{u}$ também é um múltiplo de

[4] Isso irá valer mesmo quando $\operatorname{Nuc} A = \{\mathbf{0}\}$, pois $\{\mathbf{0}\} = \operatorname{Span}\{\mathbf{0}\}$ (veja o exercício P2.18). Nesse caso, no entanto, escrever $\operatorname{Nuc} A$ dessa maneira é uma bobagem.

$\mathbf{v}$, pois $\mathbf{u} = \frac{1}{3}\mathbf{v}$ (veja o exercício P2.19(a)). Nesse sentido, os vetores $\mathbf{u}$ e $\mathbf{v}$ são "dependentes" entre si: podemos escrever um deles como uma *combinação linear* do outro.[5]

Por outro lado, os vetores $\mathbf{u} = \left[\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right]$ e $\mathbf{w} = \left[\begin{smallmatrix} 4 \\ 0 \end{smallmatrix}\right]$ são "independentes": $\mathbf{u}$ não é múltiplo de $\mathbf{w}$, nem $\mathbf{w}$ é múltiplo de $\mathbf{u}$. E quanto aos vetores $\mathbf{u}$ e $\mathbf{0}_2 = \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]$? São "dependentes" ou "independentes"? Bom, $\mathbf{u}$ *não* é um múltiplo de $\mathbf{0}_2$, mas $\mathbf{0}_2$ *é* um múltiplo de $\mathbf{u}$, pois $\mathbf{0}_2 = 0\mathbf{u}$ (veja o exercício P2.19(b)). Assim, entendemos que os vetores $\mathbf{u}$ e $\mathbf{0}_2$ são *dependentes* (afinal, é possível escrever $\mathbf{0}_2$ como a combinação linear $0\mathbf{u}$ do vetor $\mathbf{u}$, não é?).

A definição a seguir generaliza esses conceitos de "dependência" e "independência" para conjuntos contendo mais do que dois vetores.[6]

Definição 3.14
O conjunto $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$ de vetores de $\mathbb{R}^n$ é dito **linearmente independente** se o sistema linear homogêneo

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_m\mathbf{a}_m = \mathbf{0}_n \tag{3.9}$$

tem apenas a solução trivial $\mathbf{x} = \mathbf{0}_m$ (isto é, $x_1 = 0$, $x_2 = 0$, $\ldots$, $x_m = 0$).

Em contrapartida, o conjunto $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$ é dito **linearmente dependente** se o sistema (3.9) tem alguma solução não-trivial, ou seja, se existem escalares c_1, c_2, $\ldots$, c_m, *não todos iguais a zero*, tais que

$$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_m\mathbf{a}_m = \mathbf{0}_n. \tag{3.10}$$

Por simplicidade, podemos escrever "os vetores $\mathbf{a}_1$, $\ldots$, $\mathbf{a}_m$ são linearmente independentes" (ou dependentes), ao invés de "o conjunto $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ é linearmente independente" (ou dependente). Às vezes, as expressões "linearmente independente" e "linearmente dependente" são abreviadas por "LI" e "LD".

Uma equação do tipo (3.10), *quando os escalares c_j não são todos iguais a zero*, é chamada de uma **relação de dependência linear** entre os vetores $\mathbf{a}_1$, $\ldots$, $\mathbf{a}_m$. É claro que só existem relações de dependência linear entre vetores linearmente *dependentes*. Assim, $3\mathbf{u} - \mathbf{v} = \mathbf{0}_2$ e $0\mathbf{u} + 17(\mathbf{0}_2) = \mathbf{0}_2$ são relações de dependência linear entre os vetores considerados no início dessa seção (verifique!).[7]

Em uma primeira leitura, a definição 3.14 e o seu propósito podem não estar muito claros. O conceito de independência linear, no entanto, é importantíssimo, e recomendamos uma leitura atenciosa do restante dessa seção. As proposições 3.18 e 3.20, em particular, são elucidativas quanto à natureza de conjuntos linearmente dependentes e independentes. Mas vamos começar com exemplos simples, apenas para fixar ideias.

[5]Observe que uma combinação linear de um só vetor é simplesmente um múltiplo desse vetor. Assim, $\mathbf{v} = 3\mathbf{u}$ é uma combinação linear do vetor $\mathbf{u}$.

[6]Há um pequeno "deslize técnico" nessa definição. Veja a observação no final da seção.

[7]Obteríamos outras relações de dependência linear entre $\mathbf{u}$ e $\mathbf{0}_2$ substituindo o escalar 17 por qualquer outro escalar não nulo em $0\mathbf{u} + 17(\mathbf{0}_2) = \mathbf{0}_2$.

Exemplo 3.15
Sejam

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \quad \text{e} \quad \mathbf{a}_3 = \begin{bmatrix} -1 \\ 1 \\ 10 \end{bmatrix}.$$

Observe que estes são os mesmos vetores do exemplo 2.21. Vamos investigar se o conjunto $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ é linearmente independente ou não.

Temos que determinar se o sistema homogêneo $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 = \mathbf{0}$ possui unicamente a solução trivial, ou se possui uma infinidade de soluções (veja a seção 3.1). Primeiro, obtemos uma forma escalonada da matriz $\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix}$:

$$\begin{bmatrix} \boxed{1} & 1 & -1 \\ 2 & 3 & 1 \\ -1 & 2 & 10 \end{bmatrix} \xrightarrow[\ell_3 \to \ell_3 + \ell_1]{\ell_2 \to \ell_2 - 2\ell_1} \begin{bmatrix} \boxed{1} & 1 & -1 \\ 0 & \boxed{1} & 3 \\ 0 & 3 & 9 \end{bmatrix} \xrightarrow{\ell_3 \to \ell_3 - 3\ell_2} \begin{bmatrix} \boxed{1} & 1 & -1 \\ 0 & \boxed{1} & 3 \\ 0 & 0 & 0 \end{bmatrix}.$$

Agora, basta aplicar a proposição 3.3: como a matriz $\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix}$ tem uma coluna não-pivô, o sistema $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 = \mathbf{0}$ possui soluções não-triviais, e, portanto, os vetores $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{a}_3$ são linearmente *dependentes*.

Exemplo 3.16
Sejam agora

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \quad \text{e} \quad \mathbf{a}_4 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

Vejamos se o conjunto $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_4\}$ é linearmente independente.

Vamos escalonar a matriz $\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_4 \end{bmatrix}$:

$$\begin{bmatrix} \boxed{1} & 1 & 1 \\ 2 & 3 & 0 \\ -1 & 2 & 0 \end{bmatrix} \xrightarrow[\ell_3 \to \ell_3 + \ell_1]{\ell_2 \to \ell_2 - 2\ell_1} \begin{bmatrix} \boxed{1} & 1 & 1 \\ 0 & \boxed{1} & -2 \\ 0 & 3 & 1 \end{bmatrix} \xrightarrow{\ell_3 \to \ell_3 - 3\ell_2} \begin{bmatrix} \boxed{1} & 1 & 1 \\ 0 & \boxed{1} & -2 \\ 0 & 0 & \boxed{7} \end{bmatrix}.$$

Como todas as colunas da matriz $\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_4 \end{bmatrix}$ são colunas-pivô, o sistema $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_4 = \mathbf{0}$ possui apenas a solução trivial $\left[\begin{smallmatrix} x_1 \\ x_2 \\ x_3 \end{smallmatrix}\right] = \mathbf{0}$, logo o conjunto $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_4\}$ é, de fato, linearmente *independente*.

Esses exemplos motivam e ilustram o resultado a seguir.

Teorema 3.17
Seja $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m\}$ um conjunto de vetores de $\mathbb{R}^n$. Este conjunto é linearmente independente se e somente se todas as colunas da matriz $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_m \end{bmatrix}$ são colunas-pivô.

O teorema é uma consequência imediata da proposição 3.3 e da definição 3.14. De fato, ele é uma mera reformulação das proposições 3.3 e 3.11 em termos do conceito de independência linear.

A seguinte proposição fornece uma caracterização de conjuntos linearmente dependentes e, também, uma boa justificativa para o uso dessa terminologia.

Proposição 3.18
Um conjunto $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$ *de dois ou mais vetores de* $\mathbb{R}^n$ *é linearmente dependente se e somente se (pelo menos) um dos* $\mathbf{a}_k$ *é uma combinação linear dos demais vetores do conjunto.*

Demonstração: Suponha que o conjunto $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ seja linearmente dependente. Então existe alguma relação de dependência linear $c_1\mathbf{a}_1 + \cdots + c_m\mathbf{a}_m = \mathbf{0}_n$, em que os escalares c_j não são todos nulos. Assim, podemos escolher algum $c_k \neq 0$ (pode haver várias escolhas admissíveis) e reescrever essa relação como

$$c_k\mathbf{a}_k = -c_1\mathbf{a}_1 - c_2\mathbf{a}_2 - \cdots - c_{k-1}\mathbf{a}_{k-1} - c_{k+1}\mathbf{a}_{k+1} - \cdots - c_m\mathbf{a}_m,$$

ou ainda como

$$\mathbf{a}_k = -\frac{c_1}{c_k}\mathbf{a}_1 - \frac{c_2}{c_k}\mathbf{a}_2 - \cdots - \frac{c_{k-1}}{c_k}\mathbf{a}_{k-1} - \frac{c_{k+1}}{c_k}\mathbf{a}_{k+1} - \cdots - \frac{c_m}{c_k}\mathbf{a}_m.$$

A hipótese $c_k \neq 0$ foi usada nesse último passo. Se c_k fosse zero, a divisão não estaria definida! Concluímos da última equação acima que $\mathbf{a}_k$ é uma combinação linear dos demais vetores do conjunto dado.

Reciprocamente, se $\mathbf{a}_k$ é uma combinação linear dos demais vetores, então existem escalares α_j tais que

$$\mathbf{a}_k = \alpha_1\mathbf{a}_1 + \alpha_2\mathbf{a}_2 + \cdots + \alpha_{k-1}\mathbf{a}_{k-1} + \alpha_{k+1}\mathbf{a}_{k+1} + \cdots + \alpha_m\mathbf{a}_m.$$

A equação acima é equivalente a

$$\alpha_1\mathbf{a}_1 + \alpha_2\mathbf{a}_2 + \cdots + \alpha_{k-1}\mathbf{a}_{k-1} + (-1)\mathbf{a}_k + \alpha_{k+1}\mathbf{a}_{k+1} + \cdots + \alpha_m\mathbf{a}_m = \mathbf{0}_n.$$

Essa é uma relação de dependência linear entre os vetores $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$, pois os pesos na combinação linear à esquerda *não* são todos iguais a zero: o coeficiente de $\mathbf{a}_k$, ao menos, é $-1 \neq 0$. □

O corolário a seguir é, meramente, um caso particular da proposição 3.18.

Corolário 3.19
Um conjunto de dois *vetores é linearmente dependente se e somente se um dos vetores é um múltiplo do outro.*

Isso mostra que a definição 3.14 generaliza a noção intuitiva de dependência entre dois vetores que discutimos no início da seção, como havíamos anunciado. Recomendamos que você reveja aquela discussão. Mas cuidado: o corolário acima se aplica apenas a conjuntos de *dois* vetores! Veja os exercícios P3.14 e P3.15.

A próxima proposição começa a revelar por que conjuntos linearmente *independentes* são de interesse.

Na seção 2.5, vimos que um vetor $\mathbf{b}$ é uma combinação linear de $\mathbf{a}_1, \ldots, \mathbf{a}_m$ (isto é, $\mathbf{b} \in \operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$) se e somente se o sistema $x_1\mathbf{a}_1 + \cdots + x_m\mathbf{a}_m = \mathbf{b}$ é possível (ver proposição 2.16, na página 52). Não discutimos, no entanto, a questão da *unicidade* de solução desse sistema. A proposição a seguir mostra que essa questão diz respeito à independência linear do conjunto $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$.

Proposição 3.20
Seja $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$ *um conjunto de vetores de* $\mathbb{R}^n$*. Esse conjunto é linearmente independente se e somente se cada vetor* $\mathbf{b} \in \operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ *pode ser escrito* de uma única maneira *como combinação linear de* $\mathbf{a}_1, \ldots, \mathbf{a}_m$*.*

Demonstração: Suponha, primeiro, que o conjunto $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ seja linearmente independente, e seja $\mathbf{b}$ um vetor qualquer de $\operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$. Ora, se $\mathbf{b}$ pertence a $\operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$, então existem escalares $c_1, \ldots, c_m$ tais que

$$\mathbf{b} = c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_m\mathbf{a}_m. \tag{3.11}$$

Agora considere "outra" representação

$$\mathbf{b} = d_1\mathbf{a}_1 + d_2\mathbf{a}_2 + \cdots + d_m\mathbf{a}_m. \tag{3.12}$$

Vamos mostrar que $d_1 = c_1$, $d_2 = c_2$, ..., $d_m = c_m$. Dessa forma, a representação (3.11) é, na realidade, única. Subtraindo a equação (3.12) da (3.11), obtemos

$$\mathbf{0}_n = (c_1 - d_1)\mathbf{a}_1 + (c_2 - d_2)\mathbf{a}_2 + \cdots + (c_m - d_m)\mathbf{a}_m.$$

Como o conjunto $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ é linearmente independente, a equação acima implica que $c_1 - d_1 = 0$, $c_2 - d_2 = 0$, ..., $c_m - d_m = 0$, pois não há relações de dependência entre $\mathbf{a}_1, \ldots, \mathbf{a}_m$. Portanto, vale $d_j = c_j$ para $j = 1, 2, \ldots, m$.

Agora, provaremos a recíproca. Suponha que cada vetor de $\operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ tenha uma única representação como combinação linear dos vetores $\mathbf{a}_1, \ldots, \mathbf{a}_m$. O vetor zero $\mathbf{0}_n$, em particular, pertence ao *span* de $\mathbf{a}_1, \ldots, \mathbf{a}_m$ (conforme a observação 2.14, na página 51). Dessa maneira,

$$\mathbf{0}_n = 0\mathbf{a}_1 + 0\mathbf{a}_2 + \cdots + 0\mathbf{a}_m$$

é a *única* forma de representar $\mathbf{0}_n$ como combinação dos vetores $\mathbf{a}_j$. Em outras palavras, o sistema homogêneo (3.9) possui apenas a solução trivial, e, portanto, o conjunto $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ é linearmente independente. □

A proposição 3.20 será crucial na seção 4.4.

Reunimos, no teorema abaixo, diversas formas de se dizer que um conjunto é linearmente independente. Esta lista pode auxiliar o leitor, ou a leitora, em seus estudos e na resolução de exercícios.

Teorema 3.21
Sejam $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ *vetores de* $\mathbb{R}^n$*, e seja* A *a matriz de tamanho* $n \times m$ *dada por* $\begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_m\end{bmatrix}$*. As seguintes afirmativas são equivalentes:*

(a) *O conjunto* $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$ *é linearmente independente.*

(b) *Os vetores* $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ *são linearmente independentes.*

(c) *O sistema homogêneo* $x_1\mathbf{a}_1 + \cdots + x_m\mathbf{a}_m = \mathbf{0}_n$ *tem unicamente a solução trivial* $x_1 = 0$, $x_2 = 0$, ..., $x_m = 0$.

(d) *Se* $x_1\mathbf{a}_1 + \cdots + x_m\mathbf{a}_m = \mathbf{0}_n$*, então todos os* x_j *têm que ser iguais a zero.*

(e) *Não existem relações de dependência linear entre os vetores* $\mathbf{a}_1, \ldots, \mathbf{a}_m$.

(f) *Cada vetor de* $\mathrm{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ *pode ser escrito de uma única maneira como combinação linear de* $\mathbf{a}_1, \ldots, \mathbf{a}_m$.

(g) *O sistema homogêneo* $A\mathbf{x} = \mathbf{0}_n$ *tem unicamente a solução trivial* $\mathbf{x} = \mathbf{0}_m$.

(h) $\mathrm{Nuc}\, A = \{\mathbf{0}_m\}$.

(i) *O sistema* $A\mathbf{x} = \mathbf{0}_n$ *não possui "variáveis livres".*

(j) *A matriz A possui uma posição-pivô em cada uma de suas* m colunas.

A equivalência entre as afirmativas deve estar clara, tendo em vista as definições e resultados deste capítulo. Observe, em particular, que as afirmativas (f), (h) e (j) são consideradas nas proposições 3.20, 3.11 e no teorema 3.17, respectivamente. A afirmativa (d) é apenas uma outra forma de dizer (c). A afirmativa (g) também é meramente (c) escrita em forma "compacta".

Muitas vezes, a afirmativa (j) é usada, "na prática", para testar se as outras valem ou não, como fizemos nos exemplos 3.15 e 3.16.

O resultado a seguir é, essencialmente, uma reformulação da proposição 3.6.

Proposição 3.22
Se os vetores $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ *de* $\mathbb{R}^n$ *são linearmente independentes, então, necessariamente, vale* $m \leqslant n$.

Demonstração: Se $m > n$, pela proposição 3.6, o sistema $x_1\mathbf{a}_1 + \cdots + x_m\mathbf{a}_m = \mathbf{0}_n$ teria uma infinidade de soluções. Em particular, esse sistema teria soluções *não-triviais*. Assim, os vetores $\mathbf{a}_1, \ldots, \mathbf{a}_m$ não seriam independentes. □

A proposição 3.22 diz que um conjunto linearmente independente de vetores em $\mathbb{R}^n$ pode ter, *no máximo*, n elementos. Em outras palavras, um subconjunto de $\mathbb{R}^n$ que contém mais do que n vetores é, necessariamente, linearmente dependente. Um conjunto de seis vetores em $\mathbb{R}^5$, por exemplo, é "automaticamente" linearmente dependente. O mesmo vale para um conjunto de 117 vetores em $\mathbb{R}^{100}$.

A proposição 3.22 é análoga à proposição 2.25 da seção 2.7. Note, no entanto, que as direções das desigualdades são trocadas. Cuidado para não confundir esses resultados!

Observação
Rigorosamente falando, os enunciados da definição 3.14 e de diversos resultados seguintes contêm um erro. Deixamo-lo estar até aqui, a fim de não tirar o foco das ideias principais. Mas vamos, agora, remediar esse problema técnico.

Para explicar a questão, precisamos, primeiro, considerar uma sutileza da notação usual de teoria dos conjuntos. Em uma "listagem" explícita dos elementos de um conjunto, convenciona-se que elementos repetidos são irrelevantes (e podem ser ignorados). Assim, por exemplo, os conjuntos $\{a, b, a\}$ e $\{a, b\}$ são iguais: ambos representam o conjunto que contém os elementos a e b (seja lá o que forem a e b...). Em particular, se a e b forem iguais, esse conjunto na realidade conterá um só elemento: $\{a, b\} = \{a\} = \{b\}$.

Para corrigir (ou arrematar) a definição 3.14, precisamos esclarecer que ela *não respeita* essa convenção. O conjunto $\left\{\begin{bmatrix}3\\2\end{bmatrix}, \begin{bmatrix}3\\2\end{bmatrix}\right\}$, por exemplo, é linearmente *dependente*, pois $\begin{bmatrix}3\\2\end{bmatrix} - \begin{bmatrix}3\\2\end{bmatrix} = \mathbf{0}$ é uma relação de dependência entre seus elementos (há muitas outras). De acordo com a convenção de teoria dos conjuntos, no entanto, esse conjunto seria igual a $\left\{\begin{bmatrix}3\\2\end{bmatrix}\right\}$, que, por sua vez, é linearmente *independente* (verifique). Mas não há dilema algum aqui: repetimos que a definição 3.14 *não segue* a convenção de ignorar elementos repetidos em uma listagem.

Para enfatizar esse ponto, alguns autores apresentam o conjunto $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$ da definição 3.14 como um *conjunto indexado*, ou seja, um conjunto em que sejam relevantes a ordem e as repetições dos elementos. Inclusive, alguns preferem usar a notação $(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m)$ para conjuntos indexados.

Exercícios resolvidos

R3.1. Mostre que se um conjunto $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$ contém o vetor zero, então esse conjunto é, automaticamente, linearmente dependente.

Solução: Se $\mathbf{a}_j = \mathbf{0}$ (onde j é um dos índices entre 1 e m), então vale

$$0\mathbf{a}_1 + 0\mathbf{a}_2 + \cdots + 0\mathbf{a}_{j-1} + 1\mathbf{a}_j + 0\mathbf{a}_{j+1} + \cdots + 0\mathbf{a}_m = \mathbf{0}.$$

Essa é uma relação de dependência linear entre os elementos de $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ (observe que o coeficiente do vetor $\mathbf{a}_j$ não é zero), portanto, esse conjunto é linearmente dependente.

Outra forma de resolver a questão é a seguinte:

Se $\mathbf{a}_j = \mathbf{0}$, então a j-ésima coluna da matriz $\begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_m\end{bmatrix}$ é toda de zeros, portanto não pode ser coluna-pivô (justifique, pensando no efeito do processo de escalonamento sobre esta coluna). Pelo teorema 3.17, o conjunto é linearmente dependente. Esta solução é válida, mas a primeira é mais elegante. □

R3.2. Sejam $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{v}$ vetores de $\mathbb{R}^n$. Mostre que se $\mathbf{v}$ pertence a $\operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$, então o conjunto $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{v}\}$ é linearmente dependente.

Solução: A hipótese $\mathbf{v} \in \operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$ equivale a dizer que $\mathbf{v}$ é uma combinação linear de $\mathbf{a}_1$ e $\mathbf{a}_2$ (veja a proposição 2.16 da seção 2.5). Então, pela proposição 3.18, o conjunto $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{v}\}$ é linearmente dependente.

É instrutivo obter esse resultado diretamente, sem usar a proposição 3.18 (repetindo, essencialmente, a sua prova, para esse caso mais simples). Se $\mathbf{v} \in \operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$, existem escalares c_1 e c_2 tais que $\mathbf{v} = c_1\mathbf{a}_1 + c_2\mathbf{a}_2$. Isso equivale a $c_1\mathbf{a}_1 + c_2\mathbf{a}_2 - \mathbf{v} = \mathbf{0}$. Observe que esta é uma relação de dependência linear entre $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{v}$ (justifique!), portanto esses vetores são linearmente dependentes. □

R3.3. Sejam $\mathbf{a}_1, \ldots, \mathbf{a}_m$ e $\mathbf{v}$ vetores de $\mathbb{R}^n$, e suponha que $\mathcal{I} = \{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ seja um conjunto linearmente independente. Mostre que se $\mathbf{v}$ *não* pertence a $\operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$, então o conjunto $\{\mathbf{a}_1, \ldots, \mathbf{a}_m, \mathbf{v}\}$, obtido pela inclusão de $\mathbf{v}$ em $\mathcal{I}$, é, ainda, linearmente independente.

Solução: Vamos considerar a relação

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_m\mathbf{a}_m + x_{m+1}\mathbf{v} = \mathbf{0}, \tag{3.13}$$

e mostrar que todos os escalares x_j têm que ser iguais a zero (veja a afirmativa (d) do teorema 3.21). Primeiro, argumentamos que x_{m+1} é necessariamente igual a zero. De fato, se $x_{m+1} \neq 0$, poderíamos reescrever (3.13) como

$$\mathbf{v} = -\frac{x_1}{x_{m+1}}\mathbf{a}_1 - \frac{x_2}{x_{m+1}}\mathbf{a}_2 - \cdots - \frac{x_m}{x_{m+1}}\mathbf{a}_m.$$

Mas isto contradiz a hipótese de que $\mathbf{v} \notin \operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$, portanto tem que valer $x_{m+1} = 0$. Assim, podemos simplificar a relação (3.13) para

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_m\mathbf{a}_m = \mathbf{0}.$$

Como, por hipótese, os vetores $\mathbf{a}_1, \ldots, \mathbf{a}_m$ são linearmente independentes, os escalares $x_1, \ldots, x_m$ também são necessariamente iguais a zero. □

Exercícios propostos

P3.1. Verifique que os sistemas em (3.1), na página 66, são equivalentes e podem ser escritos na forma (3.2).

P3.2. Determine se os sistemas homogêneos abaixo possuem apenas a solução trivial, ou uma infinidade de soluções. Não é necessário resolver os sistemas completamente. Use a proposição 3.3 ou a 3.6.

(a) $\begin{cases} x_1 - 3x_2 = 0 \\ 2x_1 + 3x_2 = 0 \end{cases}$

(b) $x_1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \mathbf{0}$

(c) $\begin{bmatrix} 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix} \mathbf{x} = \mathbf{0}$

(d) $\begin{bmatrix} 3 & -1 & 2 \\ -6 & 2 & -4 \end{bmatrix} \mathbf{x} = \mathbf{0}$

P3.3. Considere o sistema $B\mathbf{x} = \mathbf{0}$ do exemplo 3.4. Faça o processo de escalonamento da matriz completa $\begin{bmatrix} B \mid \mathbf{0} \end{bmatrix}$. Verifique, assim, que a sua forma escalonada reduzida é aquela dada em (3.4). Observe que cada matriz obtida durante o processo tem apenas zeros na última coluna.

P3.4. Mostre que se a matriz $\left[A \mid \mathbf{0}\right]$ é linha-equivalente a $\left[G \mid \mathbf{b}\right]$, então $\mathbf{b} = \mathbf{0}$.

P3.5. (a) Mostre que se B é uma matriz escalonada (ou escalonada reduzida), então $\left[B \mid \mathbf{0}\right]$ também é (verifique os critérios da definição 1.5).
(b) Convença-se de que se A e B são linha-equivalentes, então $\left[A \mid \mathbf{0}\right]$ e $\left[B \mid \mathbf{0}\right]$ também são.
(c) Explique a seguinte afirmativa à luz do itens anteriores: "Para resolver o sistema $A\mathbf{x} = \mathbf{0}$, basta escalonar a matriz de coeficientes A."

P3.6. Verifique uma espécie de recíproca da observação 3.2: Se $\mathbf{x} = \mathbf{0}$ é uma solução do sistema $A\mathbf{x} = \mathbf{b}$, então esse sistema é homogêneo (isto é, $\mathbf{b} = \mathbf{0}$).

P3.7. Suponha que $A\mathbf{x} = \mathbf{b}$ e $B\mathbf{x} = \mathbf{d}$ sejam sistemas equivalentes, ou seja, que tenham o mesmo conjunto-solução. Mostre que $A\mathbf{x} = \mathbf{b}$ é homogêneo se e somente se $B\mathbf{x} = \mathbf{d}$ é homogêneo. *Dica*: Use o exercício anterior.

P3.8. Determine se cada afirmativa é verdadeira ou falsa. Justifique.

(a) Se a é um escalar diferente de zero e $ax = 0$, então pode-se inferir que $x = 0$.
(b) Se A é uma matriz diferente da matriz zero e $A\mathbf{x} = \mathbf{0}$, então pode-se inferir que $\mathbf{x} = \mathbf{0}$. *Observação*: Lembre que a "matriz zero" tem todas as entradas iguais a zero.
(c) Um sistema linear com cinco variáveis e três equações necessariamente tem uma infinidade de soluções.
(d) Um sistema linear homogêneo com cinco variáveis e três equações necessariamente tem uma infinidade de soluções.
(e) Um sistema linear homogêneo com três variáveis e cinco equações necessariamente tem *somente* a solução trivial.
(f) Um sistema linear homogêneo com três variáveis e cinco equações pode ter soluções não-triviais.
(g) Se um sistema homogêneo possui uma solução não-trivial, então, necessariamente, possui uma infinidade de soluções.

P3.9. Seja $G = \begin{bmatrix} 1 & 4 & -2 & 7 \\ 3 & 0 & 6 & 9 \\ -1 & 2 & -4 & -1 \end{bmatrix}$.

Determine os vetores abaixo que pertencem a $\operatorname{Nuc} G$:

$$\begin{bmatrix} 1 \\ 2 \\ 1 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 2 \\ 0 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} -1 \\ 3 \\ 2 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} -2 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{3.14}$$

P3.10. Descreva, explicitamente, o núcleo de cada matriz dada abaixo como o *span* de certos vetores. Inspire-se nos exemplos 3.12 e 3.13.

(a) $\begin{bmatrix} 2 & 0 & 4 \\ 1 & 3 & -1 \\ -1 & -2 & 0 \end{bmatrix}$.

(b) $\begin{bmatrix} 1 & 2 & 1 & -1 \\ -2 & -4 & 0 & 4 \\ 3 & 6 & 4 & -2 \end{bmatrix}$.

(c) A matriz A de (1.13), página 13. *Dica*: Sua forma escalonada reduzida já foi obtida em (1.15).

(d) A matriz G do exercício anterior.

(e) $\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$.

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

(g) $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

P3.11. Considere a matriz C do exemplo 3.5, na página 67. Faça o exercício sugerido ao final do exemplo. Observe que $\{\mathbf{0}\}$ é uma descrição explícita do núcleo de C. Escrever $\operatorname{Nuc} C = \operatorname{Span}\{\mathbf{0}\}$ também é correto, mas é bobagem.

P3.12. Considere o sistema abaixo:

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

(a) Cada equação do sistema representa uma reta em $\mathbb{R}^2$. Esboce-as.

(b) Explique por que a interseção das duas retas representa o núcleo da matriz $A = \left[\begin{smallmatrix} 3 & -1 \\ 1 & 2 \end{smallmatrix}\right]$.

(c) Olhe para o seu esboço, e conclua que $\operatorname{Nuc} A$ contém apenas a origem de $\mathbb{R}^2$.

(d) Usando a proposição 3.11, verifique diretamente que $\operatorname{Nuc} A = \{\mathbf{0}\}$.

P3.13. Em cada item, determine se os vetores dados são linearmente independentes ou dependentes.

(a) $\begin{bmatrix} 0 \\ -2 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ 7 \end{bmatrix}, \begin{bmatrix} 3 \\ -2 \\ 11 \end{bmatrix}$.

(b) $\begin{bmatrix} 0 \\ -2 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ -2 \\ 11 \end{bmatrix}$.

(c) $\begin{bmatrix} 0 \\ -2 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ -2 \\ 11 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$. *Dica:* Use a proposição 3.22.

(d) $\begin{bmatrix} 1 \\ 2 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \\ 1 \\ 1 \end{bmatrix}$.

(e) $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$. *Dica:* Veja o exercício resolvido R3.1.

(f) $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

P3.14. Verifique, "por inspeção", que os seguintes vetores são linearmente dependentes. *Dica*: Use o corolário 3.19.

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

P3.15. (a) Verifique que os seguintes vetores são linearmente dependentes:

$$\begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 3 \\ -2 \end{bmatrix}, \quad \begin{bmatrix} 4 \\ -1 \\ 0 \end{bmatrix}.$$

(b) Observe que, dentre os vetores dados, nenhum é multiplo do outro. Por que isso não contradiz o corolário 3.19?

P3.16. Prove o corolário 3.19 novamente, fazendo uso direto da definição 3.14. *Dica*: Reproduza, nesse caso mais simples, a prova da proposição 3.18.

P3.17. Seja G a matriz dada no exercício P3.9, e sejam $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$ e $\mathbf{u}_4$ seus vetores-coluna. Verifique que vale a relação de dependência linear $\mathbf{u}_1 + 2\mathbf{u}_2 + \mathbf{u}_3 - \mathbf{u}_4 = \mathbf{0}$. *Sugestão*: Aproveite o trabalho já feito no exercício P3.9. Repare, em particular, no primeiro vetor dado em (3.14).

P3.18. Cada solução não-trivial de $A\mathbf{x} = \mathbf{0}$ (caso exista) corresponde a uma relação de dependência linear entre as colunas de A. Explique.

P3.19. Se houver uma relação de dependência linear entre os vetores $\mathbf{a}_1$, $\mathbf{a}_2$, ... , $\mathbf{a}_m$, então haverá uma infinidade delas. Justifique.

P3.20. Em cada item, determine os valores de h tais que (i) $\mathbf{u} \in \operatorname{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$, e tais que (ii) o conjunto $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{u}\}$ seja linearmente dependente.

(a) $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 0 \\ 2 \\ 4 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \\ h \end{bmatrix}$.

(b) $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -2 \\ -6 \\ 2 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \\ h \end{bmatrix}$.

(c) $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -2 \\ -6 \\ 2 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 3 \\ 9 \\ h \end{bmatrix}$.

P3.21. Sejam $\mathbf{a}_1$, $\mathbf{a}_2$, ... , $\mathbf{a}_m$ e $\mathbf{v}$ vetores de $\mathbb{R}^n$. Mostre que, se $\mathbf{v}$ pertence a $\operatorname{Span}\{\mathbf{a}_1, \dots, \mathbf{a}_m\}$, então o conjunto $\{\mathbf{a}_1, \dots, \mathbf{a}_m, \mathbf{v}\}$ é linearmente dependente. Equivalentemente, se o conjunto $\{\mathbf{a}_1, \dots, \mathbf{a}_m, \mathbf{v}\}$ é linearmente independente, então $\mathbf{v}$ *não* pertence a $\operatorname{Span}\{\mathbf{a}_1, \dots, \mathbf{a}_m\}$. *Observação*: Essa é uma generalização do exercício resolvido R3.2.

P3.22. Sejam $\mathbf{a}_1$, ... , $\mathbf{a}_m$ e $\mathbf{v}$ vetores de $\mathbb{R}^n$, e suponha que $\mathbf{a}_1$, ... , $\mathbf{a}_m$ sejam linearmente independentes. Mostre que $\mathbf{v}$ *pertence* a $\operatorname{Span}\{\mathbf{a}_1, \dots, \mathbf{a}_m\}$ se e somente se o conjunto $\{\mathbf{a}_1, \dots, \mathbf{a}_m, \mathbf{v}\}$ é linearmente *dependente*. Equivalentemente, $\mathbf{v}$ *não* pertence a $\operatorname{Span}\{\mathbf{a}_1, \dots, \mathbf{a}_m\}$ se e somente se o conjunto $\{\mathbf{a}_1, \dots, \mathbf{a}_m, \mathbf{v}\}$ é linearmente *independente*.

Dica: Na realidade, o trabalho todo já foi feito nos exercícios R3.3 e P3.21. Basta juntar as peças. Note que o exercício resolvido R3.3 diz que, *sob a hipótese adicional da independência linear dos vetores* $\mathbf{a}_1$, ... , $\mathbf{a}_m$, vale a recíproca do exercício P3.21.

P3.23. Considere os vetores $\mathbf{a}_1 = \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} -2 \\ -6 \\ 2 \end{bmatrix}$ e $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$. Mostre que o conjunto $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{v}\}$ é linearmente dependente, mas que $\mathbf{v}$ *não* pertence a $\operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$. Explique por que isso não contradiz o exercício P3.22. Conclua que, em geral, *não vale* a recíproca do exercício P3.21.

P3.24. Sejam $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{v}$ vetores de $\mathbb{R}^3$, com $\mathbf{a}_1$ e $\mathbf{a}_2$ linearmente independentes.

(a) Convença-se de que $\operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$ é um *plano* passando pela origem em $\mathbb{R}^3$. *Dicas*: Os vetores $\mathbf{a}_1$ e $\mathbf{a}_2$ podem ser colineares? Algum deles pode ser nulo? Agora, relembre o exemplo 2.11.

(b) Faça dois esboços representando $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{v}$ e o plano $\operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$: um para o caso $\mathbf{v} \in \operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$, e outro para o caso $\mathbf{v} \notin \operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$. Isso dará uma interpretação geométrica, em $\mathbb{R}^3$, para o exercício P3.22.

P3.25. Interprete, geometricamente, cada item do exercício P3.20. Em cada caso, determine se $\operatorname{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$ é uma reta ou um plano em $\mathbb{R}^3$. Depois, reflita sobre o significado geométrico de $\mathbf{u} \in \operatorname{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$ e sobre o significado geométrico da dependência linear entre $\mathbf{v}_1$, $\mathbf{v}_2$ e $\mathbf{u}$.

P3.26. Determine se cada afirmativa é verdadeira ou falsa. Justifique.

(a) Se o conjunto $\{\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}\}$ é linearmente dependente, então $\mathbf{z}$ pode ser escrito como uma combinação linear dos demais vetores.

(b) Se o conjunto $\{\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}\}$ é linearmente dependente, então algum de seus elementos pode ser escrito como combinação linear dos demais.

(c) Para determinar se o conjunto $\{\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}\}$ é linearmente dependente, é uma boa ideia verificar se $\mathbf{x}$ é uma combinação linear dos demais elementos.

(d) Para determinar se o conjunto $\{\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}\}$ é linearmente dependente, é uma boa ideia verificar se $\mathbf{z}$ é uma combinação linear dos demais elementos.

(e) Se o conjunto $\{\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}\}$ é linearmente dependente, então algum de seus elementos está no subespaço gerado pelos demais.

P3.27. Sejam $\mathbf{a}_1$, $\mathbf{a}_2$, ... , $\mathbf{a}_m$ e $\mathbf{u}$ vetores de $\mathbb{R}^n$.

(a) Mostre que, se $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ for um conjunto linearmente dependente, então $\{\mathbf{a}_1, \ldots, \mathbf{a}_m, \mathbf{u}\}$ também será. Ou seja, se *acrescentarmos* vetores a um conjunto linearmente *dependente*, o resultado permanecerá dependente. *Dica*: Comece com uma relação de dependência linear entre $\mathbf{a}_1$, ... , $\mathbf{a}_m$. Como escrever, agora, uma relação de dependência linear entre os vetores $\mathbf{a}_1$, ... , $\mathbf{a}_m$ e $\mathbf{u}$?

(b) Mostre que, se $\{\mathbf{a}_1, \ldots, \mathbf{a}_m, \mathbf{u}\}$ for um conjunto linearmente independente, então $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ também será. Ou seja, se *removermos* vetores de um conjunto linearmente *independente*, o resultado permanecerá independente. *Dica*: Use o item anterior.

Capítulo 4

Subespaços, Bases e Dimensão

4.0 Introdução

Em diversos exemplos e exercícios dos capítulos anteriores, buscamos representações geométricas de certos subconjuntos de $\mathbb{R}^2$ e de $\mathbb{R}^3$. O exemplo 2.11 e os exercícios P2.16, P2.17, P2.21 e P2.22 tratam da interpretação geométrica do *span* de determinados conjuntos de vetores. Nos exemplos 3.12 e 3.13, interpretamos o núcleo de certas matrizes.

Talvez você tenha reparado que alguns tipos de subconjuntos são recorrentes nesses exemplos. Referimo-nos, em particular, a retas passando pela origem em $\mathbb{R}^2$ e $\mathbb{R}^3$ e a planos passando pela origem em $\mathbb{R}^3$. A recorrência desses tipos de subconjuntos não é um acaso: eles são exemplos de *subespaços*. Um subespaço é um tipo especial de subconjunto de $\mathbb{R}^n$, que tem papel central em álgebra linear. Veremos que o *span*, o espaço-coluna e o núcleo de uma matriz são subespaços. O conceito será definido precisamente na seção a seguir, e o restante do capítulo será dedicado a estudar suas propriedades.

Boa parte do material deste capítulo será de natureza mais abstrata do que o dos capítulos anteriores. Um pouco de abstração é inevitável nesta etapa, pois o objetivo, essencialmente, é descrever as propriedades de certos subconjuntos de $\mathbb{R}^n$, onde n pode ser qualquer inteiro positivo. Como podemos descrever um subconjunto de $\mathbb{R}^5$, de $\mathbb{R}^{17}$ ou de $\mathbb{R}^{821}$ sem usar conceitos abstratos? Uma visualização geométrica, direta e concreta desses espaços é evidentemente impossível.

4.1 Subespaços de $\mathbb{R}^n$

Comecemos fazendo uma breve recordação sobre subconjuntos. Dizer que Y é um **subconjunto** de X é o mesmo que dizer que todo elemento de Y é um elemento de X. Nesse caso, dizemos também que Y está *contido* em X, e denotamos $Y \subseteq X$.[1] Quando dizemos, então, que S é um subconjunto de $\mathbb{R}^n$,

[1] Quando x é um *elemento* do conjunto X, geralmente dizemos que x *pertence* a X, e denotamos $x \in X$. Mas podemos também dizer que x está contido em X. Cuidado com essa ambiguidade da terminologia.

queremos dizer que S é um conjunto de vetores, cada um de n coordenadas reais. Observe que, por definição, o próprio $\mathbb{R}^n$ é um subconjunto de $\mathbb{R}^n$. Por outro lado, o $\mathbb{R}^2$ *não* é um subconjunto de $\mathbb{R}^3$, pois um elemento de $\mathbb{R}^2$ (uma lista ordenada de *dois* números) *não* pertence a $\mathbb{R}^3$ (cujos elementos são listas de *três* números).

Definição 4.1
Seja V um subconjunto de $\mathbb{R}^n$. Dizemos que V é um **subespaço** de $\mathbb{R}^n$ se as seguintes condições forem satisfeitas:

1. O vetor zero $\mathbf{0}_n$ de $\mathbb{R}^n$ pertence a V.
2. Para quaisquer vetores $\mathbf{u}$ e $\mathbf{v}$ de V, a soma $\mathbf{u} + \mathbf{v}$ também pertence a V.
3. Para qualquer $\mathbf{u}$ de V e qualquer escalar $\alpha \in \mathbb{R}$, o produto $\alpha\mathbf{u}$ pertence a V.

O conjunto $\mathbb{R}^n$ é um subespaço dele próprio. De fato, $\mathbb{R}^n$ é um subconjunto dele próprio que contém o vetor zero. Além disto, se $\mathbf{u}$ e $\mathbf{v}$ são vetores de $\mathbb{R}^n$ e α é um escalar, então a soma $\mathbf{u}+\mathbf{v}$ e o produto $\alpha\mathbf{u}$ são vetores de $\mathbb{R}^n$. O conjunto $\mathbb{R}^n$, portanto, satisfaz as condições da definição 4.1.

O subconjunto $\{\mathbf{0}_n\}$ do $\mathbb{R}^n$ que contém apenas o vetor zero é também um subespaço (verifique que este conjunto satisfaz as condições da definição). Por não ter muita graça, este é chamado de **subespaço trivial** do $\mathbb{R}^n$.

Cuidado!
Atenção para a premissa da definição 4.1: Para que V seja um subespaço de $\mathbb{R}^n$, *V tem que ser um subconjunto de $\mathbb{R}^n$* (além de satisfazer as condições 1 a 3). O $\mathbb{R}^2$ *não* é um subespaço de $\mathbb{R}^3$, por exemplo, pois *não é sequer um subconjunto de $\mathbb{R}^3$* (ver o parágrafo que antecede a definição 4.1). Pela mesma razão, $\mathbb{R}^n$ *não* é um subespaço de $\mathbb{R}^m$ quando $n \neq m$.

Observação 4.2
O exercício P4.2 pede que você verifique o seguinte fato, que segue da definição 4.1. Se V é um subespaço e os vetores $\mathbf{a}_1, \ldots, \mathbf{a}_m$ pertencem a V, então qualquer combinação linear

$$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_m\mathbf{a}_m$$

também pertence a V. Ou seja, se cada $\mathbf{a}_j$ pertence a V, então o conjunto gerado por $\mathbf{a}_1, \ldots, \mathbf{a}_m$ está contido em V (em símbolos, $\text{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\} \subseteq V$).

Na realidade, vale também a recíproca. Com efeito, cada $\mathbf{a}_j$ pertence a $\text{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$, em vista da observação 2.15. Assim, se $\text{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\} \subseteq V$, então $\mathbf{a}_j \in V$. Temos, portanto, o resultado a seguir.

Proposição 4.3
Seja V um subespaço de $\mathbb{R}^n$, e sejam $\mathbf{a}_1, \ldots, \mathbf{a}_m$ vetores de $\mathbb{R}^n$. O conjunto gerado por $\mathbf{a}_1, \ldots, \mathbf{a}_m$ está contido em V ($\text{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\} \subseteq V$) se e somente se cada um dos vetores $\mathbf{a}_j$ pertence a V.

A hipótese de que V seja um subespaço é muito importante na observação 4.2 e na proposição 4.3 (veja o exercício P4.3).

Os próximos resultados mostram que o subespaço gerado por vetores, o espaço das colunas e o núcleo de uma matriz são subespaços.[2] As provas são simples e elucidativas, pois consistem na verificação direta das condições da definição 4.1.

Proposição 4.4
Sejam $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ *vetores de* $\mathbb{R}^n$. *O subespaço gerado por* $\mathbf{a}_1, \ldots, \mathbf{a}_m$ *(ou seja,* $\operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$*) é, de fato, um subespaço de* $\mathbb{R}^n$.

Demonstração: Em primeiro lugar, verificamos a premissa da definição 4.1: Como $\mathbf{a}_1, \ldots, \mathbf{a}_m$ são vetores de $\mathbb{R}^n$, é claro que suas combinações lineares também o são, logo $\operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ é um subconjunto de $\mathbb{R}^n$. Agora, a condição 1: O vetor zero $\mathbf{0}_n$ pertence a $\operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$, pela observação 2.14 (página 51). Condição 2: Se $\mathbf{u}$ e $\mathbf{v}$ pertencem a $\operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$, então existem pesos $c_1, \ldots, c_m$ e $d_1, \ldots, d_m$ tais que

$$\mathbf{u} = c_1\mathbf{a}_1 + \cdots + c_m\mathbf{a}_m \qquad \text{e} \qquad \mathbf{v} = d_1\mathbf{a}_1 + \cdots + d_m\mathbf{a}_m. \tag{4.1}$$

Assim,

$$\mathbf{u} + \mathbf{v} = (c_1 + d_1)\mathbf{a}_1 + (c_2 + d_2)\mathbf{a}_2 + \cdots + (c_m + d_m)\mathbf{a}_m,$$

isto é, $\mathbf{u} + \mathbf{v}$ é uma combinação linear dos vetores $\mathbf{a}_1, \ldots, \mathbf{a}_m$. Isso é o mesmo que dizer $\mathbf{u} + \mathbf{v} \in \operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$. Finalmente, verificamos a condição 3: Se $\mathbf{u} \in \operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$, mais uma vez podemos escrever $\mathbf{u}$ como em (4.1). E, se α é um escalar qualquer, então

$$\alpha\mathbf{u} = \alpha(c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_m\mathbf{a}_m) = (\alpha c_1)\mathbf{a}_1 + (\alpha c_2)\mathbf{a}_2 + \cdots + (\alpha c_m)\mathbf{a}_m,$$

logo $\alpha\mathbf{u} \in \operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$. $\square$

Como o espaço-coluna de uma matriz é, por definição, o espaço gerado por seus vetores-coluna, obtemos a seguinte consequência direta dessa proposição.

Corolário 4.5
Seja A uma matriz $n \times m$. O espaço-coluna de A $(\operatorname{Col} A)$ *é um subespaço de* $\mathbb{R}^n$.

Exemplo 4.6
Sejam $\mathbf{a}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$ e $\mathbf{a}_3 = \begin{bmatrix} -1 \\ 1 \\ 10 \end{bmatrix}$ os vetores do exemplo 2.21, na página 54, e seja $W = \operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$. Pela proposição 4.4, o conjunto W é um subespaço de $\mathbb{R}^3$. Mais concretamente, W é um *plano* passando pela origem de $\mathbb{R}^3$, conforme o exercício P2.21. Recapitulando esse exercício, vimos que os vetores $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{a}_3$ são *coplanares*. É fácil ver, no entanto, que esses vetores *não são* colineares. A conclusão de que o conjunto W gerado por eles seja um plano, portanto, é intuitiva (veja também o exercício P2.22). Uma região do plano W contendo os vetores $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{a}_3$, é ilustrada na figura 2.7 (página 62).

[2] Assim, a terminologia "subespaço gerado", que introduzimos no capítulo 2, está finalmente justificada.

No exemplo acima, usamos fortemente os resultados do exercício P2.21. Nos exemplos 4.28 e 4.29 da seção 4.3, veremos uma forma mais sistemática de fornecer descrições geométricas de subespaços de $\mathbb{R}^2$ e de $\mathbb{R}^3$.

Proposição 4.7
Seja A uma matriz $n \times m$. O núcleo de A ($\operatorname{Nuc} A$) é um subespaço de $\mathbb{R}^m$.

Demonstração: Primeiramente, observamos que $\operatorname{Nuc} A$ é um subconjunto de $\mathbb{R}^m$ (recorde a definição 3.8). Condição 1: Em vista da observação 3.10, o vetor zero de $\mathbb{R}^m$ pertence a $\operatorname{Nuc} A$ ($\mathbf{0}_m$ é sempre uma solução do sistema $A\mathbf{x} = \mathbf{0}_n$). Condição 2: Se $\mathbf{u}$ e $\mathbf{v}$ pertencem a $\operatorname{Nuc} A$, então, por definição, $A\mathbf{u} = \mathbf{0}_n$ e $A\mathbf{v} = \mathbf{0}_n$. Usando as propriedades algébricas do produto matriz-vetor, temos $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v} = \mathbf{0}_n + \mathbf{0}_n = \mathbf{0}_n$, logo o vetor $\mathbf{u} + \mathbf{v}$ pertence a $\operatorname{Nuc} A$. Verificamos a condição 3 similarmente: Se $\mathbf{u} \in \operatorname{Nuc} A$ e α é um escalar, então $A(\alpha\mathbf{u}) = \alpha A\mathbf{u} = \alpha\mathbf{0}_n = \mathbf{0}_n$, logo $\alpha\mathbf{u} \in \operatorname{Nuc} A$. □

Cuidado!
Mais uma vez, alertamos o leitor para não confundir $\operatorname{Col} A$ e $\operatorname{Nuc} A$. Em particular, se A é uma matriz $n \times m$, então $\operatorname{Col} A$ é um subespaço de $\mathbb{R}^n$, ao passo que $\operatorname{Nuc} A$ é um subespaço de $\mathbb{R}^m$.

4.2 Conjuntos geradores e bases de subespaços

Considere os seguintes n vetores de $\mathbb{R}^n$:

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots, \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}. \tag{4.2}$$

Qualquer vetor de $\mathbb{R}^n$ pode ser escrito como uma combinação linear desses vetores, ou seja, eles geram $\mathbb{R}^n$ (recorde a definição 2.20 e o teorema 2.24). De fato, se $\mathbf{v}$ é um vetor de coordenadas $v_1, v_2, \ldots, v_n$, então

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \cdots + v_n \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + \cdots + v_n\mathbf{e}_n. \tag{4.3}$$

Ademais, esta é a *única* forma de representar $\mathbf{v}$ como combinação linear de $\mathbf{e}_1, \ldots, \mathbf{e}_n$. Esse fato está associado à independência linear dos vetores $\mathbf{e}_1, \ldots, \mathbf{e}_n$ (conforme a proposição 3.20), mas o exercício P4.6 pede que você verifique-o diretamente.

Em síntese, o conjunto $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ tem a seguinte propriedade interessante: qualquer vetor de $\mathbb{R}^n$ pode ser escrito de uma única maneira como combinação linear de seus elementos. Dizemos que este conjunto é uma *base* de $\mathbb{R}^n$. Nesta seção vamos definir esse conceito precisamente. De fato, introduziremos o conceito de base de um *subespaço*, que não precisa ser o $\mathbb{R}^n$ inteiro.

Primeiro, precisamos generalizar a definição 2.20.

Definição 4.8
Seja $\mathcal{G} = \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$ um conjunto de vetores de $\mathbb{R}^n$ e seja V um subespaço de $\mathbb{R}^n$. Se o subespaço gerado por $\mathbf{a}_1, \ldots, \mathbf{a}_m$ coincide com V, isto é, se

$$V = \text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}, \tag{4.4}$$

dizemos que $\mathcal{G} = \{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ é um **conjunto gerador** do subespaço V. Podemos dizer também, mais informalmente, que os vetores $\mathbf{a}_1$, $\mathbf{a}_2$, ... , $\mathbf{a}_m$ **geram** V, ou ainda que o conjunto $\mathcal{G}$ **gera** V.

É claro que a condição (4.4) será válida se e somente se forem válidas as seguintes duas condições:[3] (i) $\text{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\} \subseteq V$, e (ii) $V \subseteq \text{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$. Em vista da proposição 4.3, a condição (i) equivale a dizer que cada vetor $\mathbf{a}_j$ pertence a V. A condição (ii), por sua vez, equivale a dizer que qualquer vetor de V pode ser gerado pelos vetores $\mathbf{a}_1$, ... , $\mathbf{a}_m$. Às vezes, é mais conveniente considerar separadamente estas duas "sub-condições" (como no exemplo 4.16, adiante). Sintetizamos essas considerações na seguinte observação.

Observação 4.9
O conjunto $\mathcal{G} = \{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ gera V se e somente se valem as seguintes condições: (i) cada $\mathbf{a}_j$ está em V; e (ii) qualquer vetor de V é gerado por $\mathbf{a}_1$, ... , $\mathbf{a}_m$.

Agora estamos prontos para enunciar a principal definição desta seção.

Definição 4.10
Seja V um subespaço de $\mathbb{R}^n$. Um conjunto $\mathcal{B} = \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_p\}$ de vetores de $\mathbb{R}^n$ é dito uma **base** do subespaço V se as seguintes condições estiverem satisfeitas:

1. $\mathcal{B}$ é um conjunto gerador para V, isto é, $\text{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_p\} = V$. Conforme a observação 4.9, esta condição se desdobra em duas:
   (i) Cada vetor $\mathbf{a}_j$ pertence a V.
   (ii) Qualquer vetor $\mathbf{v}$ de V pode ser gerado por $\mathbf{a}_1$, ... , $\mathbf{a}_p$.
2. O conjunto $\mathcal{B} = \{\mathbf{a}_1, \ldots, \mathbf{a}_p\}$ é linearmente independente.

Repetindo, uma base de V é um *conjunto gerador de* V que é também *linearmente independente*.

Enfatizamos a "sub-condição" 1(i): Se $\{\mathbf{a}_1, \ldots, \mathbf{a}_p\}$ é uma base de V, então *cada um dos vetores* $\mathbf{a}_j$ *tem que pertencer a* V. Na observação abaixo, destacamos outra condição necessária (mas não suficiente) para que um conjunto seja uma base de um subespaço de $\mathbb{R}^n$.

[3] É necessário que *ambas* sejam verdadeiras para que valha $V = \text{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$.

Observação 4.11
Se $\mathcal{B} = \{\mathbf{a}_1, \ldots, \mathbf{a}_p\}$ é uma base de um subespaço V de $\mathbb{R}^n$, então $\mathcal{B}$ *tem, no máximo, n elementos*. Ou seja, a condição $p \leqslant n$ é necessária. Isso segue da proposição 3.22, já que os elementos da base $\mathcal{B}$ são, afinal, vetores *linearmente independentes* de $\mathbb{R}^n$. Neste ponto, a hipótese $V \subseteq \mathbb{R}^n$ é importante.

Observação 4.12
Dizemos, *por convenção*, que a base do subespaço trivial $\{\mathbf{0}\}$ de $\mathbb{R}^n$ é o conjunto vazio $\varnothing = \{\}$. Coerentemente, dizemos também que o conjunto vazio é linearmente independente, e que é um conjunto gerador para o subespaço trivial.

Exemplo 4.13
O conjunto $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, onde os $\mathbf{e}_j$ são dados em (4.2), é uma base de $\mathbb{R}^n$. Como vimos, esse conjunto gera $\mathbb{R}^n$ e, conforme o exercício P4.7, é também linearmente independente. Esse conjunto é chamado de **base canônica** de $\mathbb{R}^n$. A base canônica do $\mathbb{R}^2$, por exemplo, é $\left\{\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]\right\}$, e a do $\mathbb{R}^3$ é $\left\{\left[\begin{smallmatrix}1\\0\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}0\\1\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}0\\0\\1\end{smallmatrix}\right]\right\}$.

Cuidado!
Estamos usando os mesmos símbolos ($\mathbf{e}_j$) para escrever vetores da base canônica de $\mathbb{R}^2$, de $\mathbb{R}^3$ ou de $\mathbb{R}^n$, para n qualquer! Denotamos, por exemplo, a base canônica de $\mathbb{R}^2$ por $\{\mathbf{e}_1, \mathbf{e}_2\}$, e a de $\mathbb{R}^3$ por $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. Mas os $\mathbf{e}_j$ do primeiro conjunto são vetores de $\mathbb{R}^2$, e os $\mathbf{e}_j$ do segundo são vetores de $\mathbb{R}^3$. Já os $\mathbf{e}_j$ de (4.2) são vetores de $\mathbb{R}^n$. Cabe a *você* atentar ao contexto para sanar essa ambiguidade.

Observação
"Base *canônica*" significa base *padrão* ou *standard*. Problemas e aplicações são muitas vezes formulados em termos da base canônica, mas às vezes esta *não* é a base mais conveniente para se usar no desenvolvimento e resolução de tais problemas. Veremos exemplos disso no capítulo 8.

Exemplo 4.14
Sejam $\mathbf{u} = \left[\begin{smallmatrix}-1\\3\end{smallmatrix}\right]$ e $\mathbf{v} = \left[\begin{smallmatrix}2\\5\end{smallmatrix}\right]$. Verifique que $\{\mathbf{u}, \mathbf{v}\}$ é uma base de $\mathbb{R}^2$.

Solução: Esse exercício é muito simples. Para verificar as condições da definição 4.10, escalonamos a matriz $\begin{bmatrix}\mathbf{u} & \mathbf{v}\end{bmatrix}$:

$$\begin{bmatrix}\mathbf{u} & \mathbf{v}\end{bmatrix} = \begin{bmatrix}\boxed{-1} & 2\\ 3 & 5\end{bmatrix} \xrightarrow{\ell_2 \to \ell_2 + 3\ell_1} \begin{bmatrix}\boxed{-1} & 2\\ 0 & \boxed{11}\end{bmatrix}.$$

Pelo teorema 2.23, como a matriz $\begin{bmatrix}\mathbf{u} & \mathbf{v}\end{bmatrix}$ tem uma posição-pivô em cada *linha*, o conjunto $\{\mathbf{u}, \mathbf{v}\}$ gera $\mathbb{R}^2$ (é evidente que $\mathbf{u}$ e $\mathbf{v}$ pertencem a $\mathbb{R}^2$, logo a condição 1(i) é satisfeita). E, pelo teorema 3.17, como a matriz tem uma posição-pivô em cada *coluna* também, esse conjunto é linearmente independente. □

Exemplo 4.15
Sejam $\mathbf{u} = \left[\begin{smallmatrix}-1\\3\end{smallmatrix}\right]$, $\mathbf{v} = \left[\begin{smallmatrix}2\\5\end{smallmatrix}\right]$ e $\mathbf{w} = \left[\begin{smallmatrix}8\\9\end{smallmatrix}\right]$. Determine se $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ é uma base de $\mathbb{R}^2$.

Solução: O conjunto $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ gera $\mathbb{R}^2$ (verifique!), mas *não* é linearmente independente (um conjunto de três vetores em $\mathbb{R}^2$ nunca o é, conforme a proposição 3.22). Assim, esse conjunto *não* é uma base de $\mathbb{R}^2$. □

Exemplo 4.16
Sejam $\mathbf{a}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$ e $\mathbf{a}_3 = \begin{bmatrix} -1 \\ 1 \\ 10 \end{bmatrix}$ os vetores do exemplo 4.6, e seja novamente $W = \operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$. Como vimos, W é um plano passando pela origem de $\mathbb{R}^3$. Evidentemente, $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ é um conjunto gerador de W, pela própria definição de W. No entanto, esse conjunto *não* é uma base de W, pois não é linearmente independente (conforme o exemplo 3.15). Vamos verificar, em contrapartida, que $\mathcal{B} = \{\mathbf{a}_1, \mathbf{a}_2\}$ é uma base de W.

A condição 1(i) da definição 4.10 é claramente satisfeita: $\mathbf{a}_1$ e $\mathbf{a}_2$ pertencem a W, em vista da observação 2.15. Para abordar a condição 1(ii), observamos primeiro que o vetor $\mathbf{a}_3$ é gerado por $\mathbf{a}_1$ e $\mathbf{a}_2$. De fato, $\mathbf{a}_3 = -4\mathbf{a}_1 + 3\mathbf{a}_2$ (verifique!). Agora, seja $\mathbf{w} = c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + c_3\mathbf{a}_3$ um vetor de W, isto é, um vetor qualquer gerado por $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{a}_3$. Ora, $\mathbf{w}$ pode ser escrito como combinação linear de $\mathbf{a}_1$ e $\mathbf{a}_2$ apenas, já que

$$\begin{aligned} \mathbf{w} &= c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + c_3\mathbf{a}_3 = \\ &= c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + c_3(-4\mathbf{a}_1 + 3\mathbf{a}_2) = (c_1 - 4c_3)\mathbf{a}_1 + (c_2 + 3c_3)\mathbf{a}_2. \end{aligned}$$

Assim, qualquer vetor de W é gerado por $\mathbf{a}_1$ e $\mathbf{a}_2$. Finalmente, temos que verificar que $\{\mathbf{a}_1, \mathbf{a}_2\}$ é um conjunto linearmente independente (condição 2). Deixamos isso como um exercício para o leitor.

Na seção 4.5, veremos um método para encontrar, no caso geral, uma base para o subespaço gerado por um dado conjunto de vetores. Abordaremos também a construção, "na prática", de bases para o núcleo e o espaço-coluna de uma dada matriz.

No momento, vamos discutir alguns resultados teóricos, considerando um subespaço qualquer de $\mathbb{R}^n$. Veremos, em particular, que todo subespaço de $\mathbb{R}^n$, de fato, possui uma base (veja o teorema 4.20, abaixo). As demonstrações são relegadas à seção 4.7, cuja leitura é opcional. No entanto, recomendamos que você leia os *enunciados* dos resultados com atenção, e procure entendê-los de uma maneira geométrica e intuitiva. Para que a discussão não fique demasiado abstrata, a construção de imagens mentais é útil. Para isso, você pode pensar em um subespaço V "abstrato" como sendo $\mathbb{R}^2$, $\mathbb{R}^3$, ou, digamos, um plano passando pela origem de $\mathbb{R}^3$.

Dito esse preâmbulo, passemos aos resultados. Vimos no exemplo 4.16, acima, que o vetor $\mathbf{a}_3$ é, em certo sentido, um elemento "redundante" do conjunto gerador $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ de W. Quando retiramos esse vetor, o conjunto $\{\mathbf{a}_1, \mathbf{a}_2\}$ resultante permanece, ainda, um conjunto gerador de W. Isso se deve ao fato de que $\mathbf{a}_3$ é, ele próprio, gerado por $\mathbf{a}_1$ e $\mathbf{a}_2$. O lema a seguir generaliza esse resultado.

Lema 4.17
Seja V um subespaço de $\mathbb{R}^n$, e seja $\mathcal{G} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m\}$ um conjunto gerador de V. Se um dos vetores de $\mathcal{G}$ — digamos, $\mathbf{a}_k$ — é uma combinação linear dos demais, então o conjunto $\mathcal{G}' = \{\mathbf{a}_1, \dots, \mathbf{a}_{k-1}, \mathbf{a}_{k+1}, \dots, \mathbf{a}_m\}$, obtido pela remoção do vetor $\mathbf{a}_k$ de $\mathcal{G}$, ainda é um conjunto gerador de V.

Observação
Já poderíamos ter enunciado e provado a essência desse lema no capítulo 2 (veja o exercício P4.8). Postergamo-lo até aqui para que pudéssemos usar a linguagem da definição 4.8 e para que estivéssemos mais habituados aos conceitos envolvidos.

Teorema 4.18
Se $\mathcal{G} = \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$ é um conjunto gerador do subespaço V, então algum subconjunto de $\mathcal{G}$ (possivelmente o próprio $\mathcal{G}$) é uma base de V. Em outras palavras, ou $\mathcal{G}$ é uma base de V, ou então podemos obter uma base de V pelo processo de remover de $\mathcal{G}$ um ou mais elementos.

O teorema acima diz que qualquer conjunto gerador de V, se já não for uma base de V, pode ser "reduzido" a uma base, mediante a exclusão de vetores "redundantes". Um caso particular simples foi ilustrado no exemplo 4.16. Veja também o exercício P4.9.

O teorema a seguir é análogo ao anterior. Ele diz que qualquer conjunto linearmente independente de vetores de V, se já não for uma base de V, pode ser "complementado" para formar uma base. O ingrediente principal da demonstração é o exercício resolvido R3.3 (página 77).

Teorema 4.19
Seja V um subespaço de $\mathbb{R}^n$. Se $\mathcal{I}$ é um conjunto linearmente independente de vetores de V, então existe uma base de V que contém o conjunto $\mathcal{I}$. Em outras palavras, ou $\mathcal{I}$ é uma base de V, ou então podemos obter uma base pelo processo de acrescentar a $\mathcal{I}$, apropriadamente, um ou mais vetores de V.

O teorema acima tem a importante consequência a seguir.

Teorema 4.20
Todo subespaço de $\mathbb{R}^n$ possui uma base finita.

Demonstração: Seja V um subespaço qualquer de $\mathbb{R}^n$. O conjunto vazio $\varnothing$ é linearmente independente (pela observação 4.12) e está contido em V.[4] Assim, podemos aplicar o teorema 4.19 e acrescentar vetores ao conjunto vazio $\varnothing$ de maneira a obter uma base de V.[5] $\square$

Perceba que a definição 4.10 requer, tacitamente, que uma base de um subespaço de $\mathbb{R}^n$ seja composta por um número *finito* de vetores.[6] Assim, no contexto de subespaços de $\mathbb{R}^n$, a expressão *base finita* é redundante. Às vezes, no entanto, é importante enfatizar a finitude, como fizemos no enunciado do teorema 4.20.

[4]Ou seja, todo elemento de $\varnothing$ é um elemento de V. Por estranha que pareça, essa afirmativa vale por vacuidade. Afinal, não existem elementos de $\varnothing$ que *não* pertençam a V (pois, simplesmente, não existem elementos de $\varnothing$!). De fato, o conjunto vazio é um subconjunto de qualquer conjunto, como você talvez já saiba.

[5]Se você não estiver convencido de que esse argumento seja válido, verifique que a prova do teorema 4.19, na seção 4.7, funciona *mesmo no caso em que o conjunto $\mathcal{I}$ é vazio.* Basta substituir as referências a $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ por referências a $\varnothing = \{\}$, o que corresponde ao caso $m = 0$. Você terá que adaptar, também, o exercício R3.3.

[6]Mesmo que a finitude não fosse imposta por definição, ela ainda estaria garantida pela observação 4.11.

Esse teorema é válido mesmo no caso do subespaço trivial $\{\mathbf{0}\}$ de $\mathbb{R}^n$. Com efeito, o subespaço trivial tem uma base finita: o conjunto vazio, que tem *zero* elementos (veja a observação 4.12).

O teorema 4.20 garante a *existência* de uma base para qualquer subespaço de $\mathbb{R}^n$, mas nada afirma sobre a *unicidade*. A base canônica $\{\mathbf{e}_1, \mathbf{e}_2\}$ e o conjunto $\{\mathbf{u}, \mathbf{v}\}$ do exemplo 4.14 são ambos bases de $\mathbb{R}^2$. Isso mostra que um subespaço pode ter mais de uma base. Na realidade, um subespaço que não seja o trivial *sempre* possui uma *infinidade* de bases, cada uma das quais composta por um número *finito* de vetores (veja o exercício P4.12). Em contrapartida, o conjunto vazio é a única base do subespaço trivial.

Observação
A proposição 4.4 diz que o *span* de um dado conjunto de vetores é um subespaço. Segue do teorema 4.20 que *todo* subespaço V de $\mathbb{R}^n$ pode ser descrito como o *span* de certos vetores. Com efeito, se $\mathcal{B} = \{\mathbf{a}_1, \ldots, \mathbf{a}_p\}$ é uma base de V, então $V = \operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_p\} = \operatorname{Col}\begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_m\end{bmatrix}$. Isso significa que o *span* e o espaço-coluna não são "tipos especiais" de subespaços. É mais correto pensar no *span* e no espaço-coluna como *formas de descrever* subespaços. Também é verdade que qualquer subespaço de $\mathbb{R}^n$ pode ser descrito como o núcleo de certa matriz (assim, o núcleo também não é um "exemplo especial" de subespaço), mas isso não será provado aqui.

4.3 Dimensão de um subespaço

Discutimos o conceito de *dimensão*, informalmente, no capítulo 2 (subseção 2.1.2). Nesta seção, finalmente, estudaremos o conceito precisamente. Uma boa internalização da ideia de dimensão, em nível "intuitivo", é valiosa. Ela auxilia muito na compreensão da natureza dos subespaços. Assim, recomendamos uma leitura cuidadosa desta seção, ainda que você decida omitir as demonstrações.

O lema a seguir é a peça fundamental desta seção, da qual decorrem naturalmente os demais resultados. Sua demonstração, no entanto, é relegada à seção 4.7.

Lema 4.21
Seja V um subespaço de $\mathbb{R}^n$, e suponha que V tenha um conjunto gerador com m elementos. Então qualquer conjunto contendo mais do que m vetores de V é, necessariamente, linearmente dependente.

A parte (a) da proposição abaixo generaliza a proposição 3.22 (página 75). A parte (b) generaliza a proposição 2.25 (página 57).

Proposição 4.22
Seja V um subespaço de $\mathbb{R}^n$, e suponha que V tenha uma base contendo p elementos. Então valem as seguintes afirmativas.

(a) *Um conjunto contendo mais do que p vetores de V é linearmente dependente.*

(b) *Um conjunto contendo menos do que p vetores de V não gera V.*

Demonstração: Seja $\mathcal{B}$ uma base de V contendo p vetores. Por ser uma base, $\mathcal{B}$ é um conjunto gerador de V. Assim, a afirmativa (a) segue da aplicação direta do lema 4.21 (com p no papel de m). Agora, se existisse um conjunto gerador de V com menos do que p elementos, o conjunto $\mathcal{B}$ teria que ser linearmente dependente, em razão, novamente, do lema. Mas $\mathcal{B}$ é linearmente independente, por ser uma base. Ou seja, a *negação* da afirmativa (b) leva a uma contradição com as hipóteses da proposição. Conclui-se que (b) tem que ser verdadeira. □

O teorema 4.20 garante a existência de uma base para qualquer subespaço de $\mathbb{R}^n$, mas, além da finitude, nada diz a respeito da *quantidade* de elementos em uma base. O teorema a seguir, que é crucial, aborda essa questão.

Teorema 4.23
Seja V um subespaço de $\mathbb{R}^n$. Se V possuir uma base composta por p vetores, então qualquer base de V conterá exatamente p vetores. Em outras palavras, dado um subespaço de $\mathbb{R}^n$, qualquer uma de suas bases contém o mesmo número de elementos.

Demonstração: Suponha que V tenha uma base contendo p elementos. Se um conjunto $\mathcal{S}$ de vetores de V contiver *mais* do que p vetores, então ele será linearmente dependente, pela afirmativa (a) da proposição 4.22. Se $\mathcal{S}$ contiver *menos* do que p vetores, ele não poderá gerar V, pela afirmativa (b). Em qualquer um dos casos, $\mathcal{S}$ não será uma base de V. Portanto, para ser uma base de V, um conjunto deverá conter exatamente p elementos. □

Agora, estamos prontos para definir o conceito de dimensão.

Definição 4.24
Seja V um subespaço de $\mathbb{R}^n$. A **dimensão** de V, denotada por $\dim V$, é o número de vetores contidos em qualquer base de V.

Essa definição tem sentido graças aos teoremas 4.20 e 4.23. Se um subespaço de $\mathbb{R}^n$ não tivesse base alguma, qual seria sua dimensão? E se, digamos, um subespaço tivesse uma base com quatro vetores e outra com cinco? Sua dimensão seria quatro ou seria cinco? Ou seria outro número? Os teoremas citados garantem que estes dilemas nunca ocorrem. Dado um subespaço, bases existem e todas elas têm igual número de vetores. Este número, que chamamos de dimensão, está, portanto, bem definido.

Exemplo 4.25 (A dimensão de $\mathbb{R}^n$)
A base canônica de $\mathbb{R}^n$ tem n vetores (veja o exemplo 4.13). De acordo com o teorema 4.23, *qualquer* base de $\mathbb{R}^n$ terá precisamente n vetores. Assim, a dimensão de $\mathbb{R}^n$ é igual a n, conforme a definição 4.24. Observe que essa conclusão condiz com a convenção que adotamos na subseção 2.1.2.

A dimensão do subespaço trivial $\{\mathbf{0}\}$ é igual a zero. Isso decorre da convenção de que o conjunto vazio $\varnothing$, que contém zero vetores, é uma base de $\{\mathbf{0}\}$.

Lembre que um conjunto $\mathcal{S}$ de vetores será uma base de V se for *linearmente independente* e *gerar* V. O teorema abaixo diz que, caso saibamos que $\mathcal{S}$

tem a quantidade "certa" de vetores, só precisamos verificar *uma* dessas duas propriedades. Para isso, no entanto, é preciso conhecer *a priori* a dimensão de V.

Teorema 4.26
Seja V um subespaço de $\mathbb{R}^n$ de dimensão igual a p.

(a) *Se $\mathcal{S}$ é um conjunto linearmente independente contendo exatamente p vetores de V, então $\mathcal{S}$ "automaticamente" gera V, e, portanto, é uma base de V.*

(b) *Se $\mathcal{S}$ é um conjunto gerador de V contendo exatamente p vetores, então $\mathcal{S}$ é "automaticamente" linearmente independente, e, portanto, é uma base de V.*

Esse teorema é uma consequência direta dos resultados desta seção e da anterior. A demonstração detalhada está na seção 4.7.

Observação 4.27
O teorema 4.26 é muito útil. Ele implica, em particular, no seguinte fato: *um conjunto linearmente independente contendo n vetores de $\mathbb{R}^n$ é, automaticamente, uma base de* $\mathbb{R}^n$. Basta lembrar que a dimensão de $\mathbb{R}^n$ é n, e aplicar a parte (a) do teorema. Similarmente, *um conjunto gerador de $\mathbb{R}^n$ com n elementos é uma base de* $\mathbb{R}^n$ (basta aplicar a parte (b)).

Podemos usar a dimensão para classificar os subespaços de $\mathbb{R}^n$. Nos exemplos a seguir, consideramos os subespaços de $\mathbb{R}^2$ e de $\mathbb{R}^3$, cujas classificações possuem interpretações geométricas simples.[7]

Exemplo 4.28 (Classificação dos subespaços de $\mathbb{R}^2$)
Um subespaço de $\mathbb{R}^2$ tem dimensão, no máximo, igual a 2. Isso é uma consequência da observação 4.11: uma base de um subespaço de $\mathbb{R}^2$ tem, no máximo, dois elementos. Assim, um subespaço de $\mathbb{R}^2$ tem dimensão igual a 0, 1 ou 2. Consideremos cada caso:

- *Subespaços de dimensão zero*: O único subespaço de dimensão zero de $\mathbb{R}^2$ é o subespaço trivial $\{\mathbf{0}_2\}$. Geometricamente, esse subespaço é representado por um ponto: a origem de $\mathbb{R}^2$.

- *Subespaços unidimensionais*: São os subespaços gerados por um único vetor não-nulo. Geometricamente, esses subespaços são as retas em $\mathbb{R}^2$ que passam pela origem.

- *Subespaços bidimensionais*: O único subespaço de $\mathbb{R}^2$ de dimensão igual a dois é o próprio $\mathbb{R}^2$. De fato, um subespaço de dimensão dois é gerado por dois vetores linearmente independentes (pois tem uma base contendo dois elementos). Mas dois vetores linearmente independentes em $\mathbb{R}^2$ geram o $\mathbb{R}^2$ todo, pela observação 4.27. Geometricamente, o $\mathbb{R}^2$ é representado por um plano, como já sabemos.

[7] Os exemplos mostrarão que o conceito de dimensão ajuda na "classificação geométrica" dos subespaços de $\mathbb{R}^n$. Essa interpretação geométrica, por sua vez, ajuda muito na assimilação do conceito de dimensão. Do ponto de vista estritamente lógico, isso pode parecer um círculo vicioso. Didaticamente, no entanto, é um círculo *virtuoso*. Tire proveito disso!

Exemplo 4.29 (Classificação dos subespaços de $\mathbb{R}^3$)
Um subespaço de $\mathbb{R}^3$ tem dimensão igual a 0, 1, 2 ou 3 (justifique!). Consideremos cada caso:

- *Subespaços de dimensão zero*: Somente o subespaço trivial $\{\mathbf{0}_3\}$. Geometricamente, esse subespaço é representado por um ponto: a origem de $\mathbb{R}^3$.
- *Subespaços unidimensionais*: São os subespaços gerados por um único vetor não-nulo. Geometricamente, esses subespaços são as retas em $\mathbb{R}^3$ que passam pela origem.
- *Subespaços bidimensionais*: São os subespaços gerados por dois vetores linearmente independentes. Geometricamente, esses subespaços são os planos em $\mathbb{R}^3$ que passam pela origem.
- *Subespaços tridimensionais*: O único subespaço de $\mathbb{R}^3$ de dimensão igual a três é o próprio $\mathbb{R}^3$. O argumento é análogo ao do exemplo anterior, e usa o fato de que três vetores linearmente independentes em $\mathbb{R}^3$ geram o $\mathbb{R}^3$ todo.

Podemos generalizar esses exemplos (veja o exercício P4.13), e concluir que existem exatamente $n+1$ tipos de subespaços de $\mathbb{R}^n$: o de dimensão zero (apenas o subespaço trivial $\{\mathbf{0}_n\}$), os de dimensão um, os de dimensão dois, e assim por diante, até o de dimensão igual a n (apenas o próprio $\mathbb{R}^n$).

4.4 Coordenadas com respeito a uma base

No início da seção 4.2, vimos que qualquer vetor de $\mathbb{R}^n$ pode ser representado de uma única maneira como combinação linear dos vetores da base canônica $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$. O teorema a seguir mostra que essa é uma propriedade de qualquer base, em qualquer subespaço.

Teorema 4.30
Seja $\mathcal{B} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_p\}$ uma base para um subespaço V de $\mathbb{R}^n$. Para cada vetor $\mathbf{v}$ de V, existe uma única lista de escalares $c_1, \dots, c_p$ tal que

$$\mathbf{v} = c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_p\mathbf{a}_p. \tag{4.5}$$

Ou seja, existe uma única maneira de expressar $\mathbf{v} \in V$ como combinação linear dos elementos de $\mathcal{B}$.

Demonstração: O conjunto $\mathcal{B}$ gera V, porque é, por hipótese, uma base de V. Assim, para cada $\mathbf{v} \in V$, existem escalares $c_1, \dots, c_p$ satisfazendo (4.5). Como o conjunto $\mathcal{B}$ é também linearmente independente, essa representação (4.5) é única, em vista da proposição 3.20. □

Exemplo 4.31
Sejam $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{a}_3$ os vetores do exemplo 4.16. Já estabelecemos que $\mathcal{B} = \{\mathbf{a}_1, \mathbf{a}_2\}$ é

uma base de $W = \operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$. Considere agora o vetor $\mathbf{v} = \begin{bmatrix} 3 \\ 4 \\ -9 \end{bmatrix}$. Observe que $\mathbf{v}$ é uma combinação linear de $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{a}_3$ (isto é, $\mathbf{v} \in W$), pois

$$\mathbf{a}_1 + \mathbf{a}_2 - \mathbf{a}_3 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \\ 10 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ -9 \end{bmatrix} = \mathbf{v}.$$

No entanto, existem muitas maneiras de escrever $\mathbf{v}$ como combinação linear de $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{a}_3$, já que estes vetores são linearmente dependentes (veja o exemplo 3.15 e a proposição 3.20). Vale também, por exemplo,

$$9\mathbf{a}_1 - 5\mathbf{a}_2 + \mathbf{a}_3 = 9\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} - 5\begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \\ 10 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ -9 \end{bmatrix} = \mathbf{v}.$$

Para se obter *todas* as maneiras de escrever $\mathbf{v}$ como combinação linear de $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{a}_3$, é necessário resolver o sistema $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 = \mathbf{v}$.

Em contrapartida, pelo teorema 4.30, $\mathbf{v} = 5\mathbf{a}_1 - 2\mathbf{a}_2$ é a *única* forma de representar $\mathbf{v}$ como combinação linear de $\mathbf{a}_1$ e $\mathbf{a}_2$, já que $\mathcal{B} = \{\mathbf{a}_1, \mathbf{a}_2\}$ é uma base de W. Você pode verificar, usando a teoria do capítulo 1 diretamente, que $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \end{bmatrix}$ é a única solução do sistema $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = \mathbf{v}$.

Neste exemplo, usamos o vetor $\mathbf{v}$ em particular, mas resultados similares valem para qualquer vetor de W.

Vamos tecer algumas conclusões a respeito do exemplo acima. Vimos, no exemplo 4.16, que os conjuntos $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ e $\{\mathbf{a}_1, \mathbf{a}_2\}$ ambos geram W (em símbolos, $W = \operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\} = \operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$).[8] Assim, podemos representar vetores de W usando combinações lineares de $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{a}_3$, ou de $\mathbf{a}_1$ e $\mathbf{a}_2$ apenas. Mas as representações usando apenas $\mathbf{a}_1$ e $\mathbf{a}_2$ são claramente mais *econômicas*, pois envolvem apenas dois vetores (compare a representação $\mathbf{v} = 5\mathbf{a}_1 - 2\mathbf{a}_2$ com $\mathbf{v} = \mathbf{a}_1 + \mathbf{a}_2 - \mathbf{a}_3$ ou com $\mathbf{v} = 9\mathbf{a}_1 - 5\mathbf{a}_2 + \mathbf{a}_3$). Além disso, tais representações são *únicas*, inambíguas, visto que $\{\mathbf{a}_1, \mathbf{a}_2\}$ é uma base de W. Em síntese, é vantajoso representar os vetores de W usando uma base. Isso motiva a definição abaixo.

Definição 4.32
Seja V um subespaço de $\mathbb{R}^n$, seja $\mathcal{B} = \{\mathbf{a}_1, \dots, \mathbf{a}_p\}$ uma base de V, e seja $\mathbf{v}$ um vetor de V. Pelo teorema 4.30, existe uma única representação

$$\mathbf{v} = c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_p\mathbf{a}_p \tag{4.6}$$

de $\mathbf{v}$ como combinação linear de $\mathbf{a}_1, \dots, \mathbf{a}_p$. Nestas condições, os escalares $c_1, \dots, c_p$ são chamados de **coordenadas do vetor $\mathbf{v}$ com relação à base $\mathcal{B}$**. O vetor de $\mathbb{R}^p$ cujas coordenadas são $c_1, \dots, c_p$ é chamado de **vetor de coordenadas de $\mathbf{v}$ com relação a $\mathcal{B}$**, que denotamos por

$$[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix}.$$

[8] Lembre-se de que W foi *definido* como $\operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ no exemplo 4.16.

Atente para as premissas da definição acima. Enfatizamos que só faz sentido falar de "coordenadas de $\mathbf{v}$ com respeito a $\mathcal{B}$" se $\mathcal{B}$ for uma *base* de um subespaço, e se $\mathbf{v}$ for um vetor pertencente a *esse mesmo subespaço*.

Um vetor de V fica completamente determinado (ou "especificado") por meio de suas coordenadas com relação a uma dada base de V. De fato, se $c_1, \ldots, c_p$ são as coordenadas de $\mathbf{v}$ com respeito à base $\{\mathbf{a}_1, \ldots, \mathbf{a}_p\}$, então o vetor $\mathbf{v}$ fica determinado pela relação (4.6) (veja o exercício resolvido R4.1).

Observação 4.33
Na definição 4.32, o vetor $\mathbf{v}$ pertence a $\mathbb{R}^n$, pois $\mathbf{v}$ é um vetor do subespaço V, e, por hipótese, $V \subseteq \mathbb{R}^n$. Repare, no entanto, que $[\mathbf{v}]_\mathcal{B}$ é um vetor de $\mathbb{R}^p$! Perceba também que p é a dimensão de V (já que é o número de elementos na base $\mathcal{B}$). Veja o exemplo 4.37, adiante.

Encontrar as coordenadas de um vetor $\mathbf{v} \in V$ com relação a uma base $\{\mathbf{a}_1, \ldots, \mathbf{a}_p\}$ de V é simples. Basta resolver o sistema $x_1\mathbf{a}_1 + \cdots + x_p\mathbf{a}_p = \mathbf{v}$.

Exemplo 4.34
Sejam $\mathbf{u} = \left[\begin{smallmatrix}-1\\3\end{smallmatrix}\right]$, $\mathbf{v} = \left[\begin{smallmatrix}2\\5\end{smallmatrix}\right]$ e $\mathbf{w} = \left[\begin{smallmatrix}8\\9\end{smallmatrix}\right]$ os vetores dos exemplos 4.14 e 4.15. Já vimos que $\{\mathbf{u}, \mathbf{v}\}$ é uma base de $\mathbb{R}^2$. Determine as coordenadas de $\mathbf{w} \in \mathbb{R}^2$ com respeito a essa base.

Solução: Temos que resolver o sistema $x_1\mathbf{u} + x_2\mathbf{v} = \mathbf{w}$, para obter a (única) representação de $\mathbf{w}$ como combinação linear de $\mathbf{u}$ e $\mathbf{v}$. Escalonando a matriz completa desse sistema, obtemos:

$$[\mathbf{u}\ \ \mathbf{v} \mid \mathbf{w}] = \left[\begin{array}{cc|c} \boxed{-1} & 2 & 8 \\ 3 & 5 & 9 \end{array}\right] \xrightarrow{\ell_2 \to \ell_2 + 3\ell_1} \left[\begin{array}{cc|c} \boxed{-1} & 2 & 8 \\ 0 & \boxed{11} & 33 \end{array}\right] \longrightarrow$$

$$\xrightarrow[\ell_2 \to \ell_2/11]{\ell_1 \to -\ell_1} \left[\begin{array}{cc|c} \boxed{1} & -2 & -8 \\ 0 & \boxed{1} & 3 \end{array}\right] \xrightarrow{\ell_1 \to \ell_1 + 2\ell_2} \left[\begin{array}{cc|c} \boxed{1} & 0 & -2 \\ 0 & \boxed{1} & 3 \end{array}\right].$$

A única solução do sistema é $x_1 = -2$, $x_2 = 3$ e temos, portanto, $\mathbf{w} = -2\mathbf{u} + 3\mathbf{v}$. Isso significa que as coordenadas de $\mathbf{w}$ com respeito à base $\{\mathbf{u}, \mathbf{v}\}$ são -2 e 3, e o vetor de coordenadas de $\mathbf{w}$ com respeito a essa base é $[\mathbf{w}]_{\{\mathbf{u},\mathbf{v}\}} = \left[\begin{smallmatrix}-2\\3\end{smallmatrix}\right]$. □

Exemplo 4.35
Sejam agora $\mathbf{y} = \left[\begin{smallmatrix}2\\3\end{smallmatrix}\right]$, $\mathbf{z} = \left[\begin{smallmatrix}2\\0\end{smallmatrix}\right]$. Verifique que $\{\mathbf{y}, \mathbf{z}\}$ é uma base de $\mathbb{R}^2$, e calcule as coordenadas do vetor $\mathbf{w} = \left[\begin{smallmatrix}8\\9\end{smallmatrix}\right]$ com relação a essa base.

Solução: Deixamos a verificação de que $\{\mathbf{y}, \mathbf{z}\}$ é uma base como um exercício para o leitor. Para calcular as coordenadas de $\mathbf{w}$ com respeito a essa base, procedemos como no exemplo anterior, resolvendo o sistema $x_1\mathbf{y} + x_2\mathbf{z} = \mathbf{w}$:

$$[\mathbf{y}\ \ \mathbf{z} \mid \mathbf{w}] = \left[\begin{array}{cc|c} 2 & 2 & 8 \\ 3 & 0 & 9 \end{array}\right] \xrightarrow[\ell_2 \to \ell_2/3]{\ell_1 \to \ell_1/2} \left[\begin{array}{cc|c} 1 & 1 & 4 \\ \boxed{1} & 0 & 3 \end{array}\right] \longrightarrow$$

$$\xrightarrow{\ell_1 \leftrightarrow \ell_2} \left[\begin{array}{cc|c} \boxed{1} & 0 & 3 \\ 1 & 1 & 4 \end{array}\right] \xrightarrow{\ell_2 \to \ell_2 - \ell_1} \left[\begin{array}{cc|c} \boxed{1} & 0 & 3 \\ 0 & \boxed{1} & 1 \end{array}\right].$$

Assim, temos $\mathbf{w} = 3\mathbf{y} + \mathbf{z}$ e, portanto, o vetor de coordenadas de $\mathbf{w}$ com respeito à base $\{\mathbf{y}, \mathbf{z}\}$ é $[\mathbf{w}]_{\{\mathbf{y},\mathbf{z}\}} = \left[\begin{smallmatrix}3\\1\end{smallmatrix}\right]$. □

Observe que o vetor $\mathbf{w}$ é o *mesmo* nos dois exemplos anteriores (a saber, é o vetor $\left[\begin{smallmatrix}8\\9\end{smallmatrix}\right] \in \mathbb{R}^2$). No entanto, seus vetores de coordenadas $[\mathbf{w}]_{\{\mathbf{u},\mathbf{v}\}}$ e $[\mathbf{w}]_{\{\mathbf{y},\mathbf{z}\}}$ com relação a bases distintas são também distintos. É comum dizermos que uma base de um subespaço fornece um *sistema de coordenadas* nesse subespaço. Bases distintas fornecem sistemas de coordenadas distintos. Isso é ilustrado, usando os exemplos acima, pela figura 4.1.

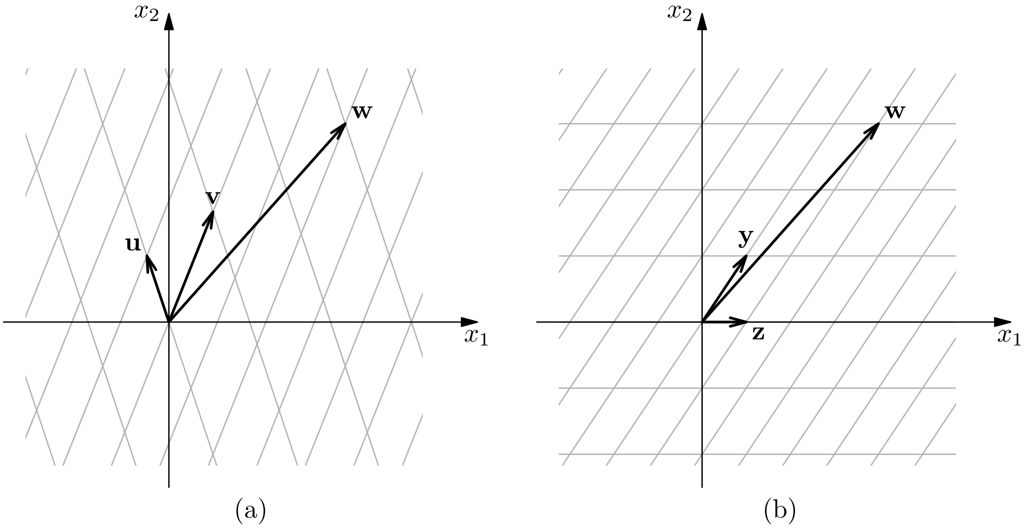


Figura 4.1: Dois sistemas de coordenadas em $\mathbb{R}^2$.

O reticulado na figura 4.1(a) é uma representação gráfica do sistema de coordenadas em $\mathbb{R}^2$ induzido pela base $\{\mathbf{u}, \mathbf{v}\}$. Os pontos onde as linhas se encontram representam as combinações lineares de $\mathbf{u}$ e $\mathbf{v}$ com pesos inteiros. Verifique geometricamente, aplicando a regra do paralelogramo diretamente sobre a figura, que $\mathbf{w} = -2\mathbf{u} + 3\mathbf{v}$, conforme o exemplo 4.34.

Já o reticulado na figura 4.1(b) representa o sistema de coordenadas associado à base $\{\mathbf{y}, \mathbf{z}\}$. Usando a regra do paralelogramo, mais uma vez, verifique que $\mathbf{w} = 3\mathbf{y} + \mathbf{z}$, conforme o exemplo 4.35. Observe, novamente, que o vetor $\mathbf{w}$ é o *mesmo* nas duas figuras (as setas que o representam são, de fato, idênticas). O que mudou da figura (a) para a (b) foi somente o sistema de coordenadas considerado.

A base canônica também fornece um sistema de coordenadas, que é simplesmente o sistema "usual" de coordenadas cartesianas. A figura 2.1(a) (página 36) ilustra o sistema de coordenadas cartesianas em $\mathbb{R}^2$. Nas figuras 4.1(a) e (b), os eixos x_1 e x_2 são apenas uma "lembrança" do sistema cartesiano (esses eixos não são adequados aos sistemas de coordenadas considerados nas figuras).

Exemplo 4.36
Seja $\mathcal{C} = \{\mathbf{e}_1, \mathbf{e}_2\}$ a base canônica do $\mathbb{R}^2$, e seja novamente $\mathbf{w} = \left[\begin{smallmatrix}8\\9\end{smallmatrix}\right]$. É evidente que $\mathbf{w} = 8\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right] + 9\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right] = 8\mathbf{e}_1 + 9\mathbf{e}_2$. Assim, as coordenadas de $\mathbf{w}$ com relação à base canônica são 8 e 9.

Mais geralmente, se $\mathbf{v}$ é um vetor de $\mathbb{R}^n$ de coordenadas $v_1, v_2, \ldots, v_n$, então a equação (4.3) (página 86) mostra que $v_1, \ldots, v_n$ são as coordenadas de $\mathbf{v}$ *com* *respeito à base canônica de* $\mathbb{R}^n$. Ou seja, quando nos referimos simplesmente às "coordenadas" de um vetor de $\mathbb{R}^n$, sem referência a uma base (como vínhamos fazendo até esta seção), a qualificação adicional "com respeito à base canônica" fica subentendida.

O próximo exemplo é mais interessante do que os anteriores, pois descreve um sistema de coordenadas em um subespaço *próprio* de $\mathbb{R}^3$ (isto é, em um subespaço que não é o $\mathbb{R}^3$ inteiro).

Exemplo 4.37
Sejam $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{a}_3$ os vetores dos exemplos 4.6 e 4.16. Já verificamos, no segundo desses exemplos, que $\mathcal{B} = \{\mathbf{a}_1, \mathbf{a}_2\}$ é uma base do plano $W = \text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$. A base $\mathcal{B}$, portanto, fornece um sistema de coordenadas em W, conforme ilustra a figura 4.2 (compare com a figura 2.7 da página 62).

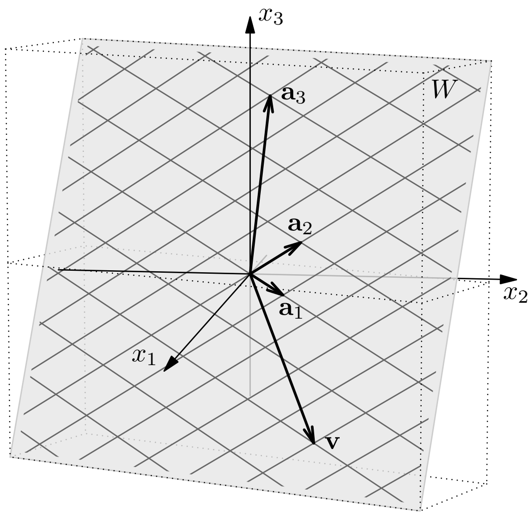


Figura 4.2: O sistema de coordenadas no plano W dado pela base $\mathcal{B} = \{\mathbf{a}_1, \mathbf{a}_2\}$.

O vetor $\mathbf{v} = \begin{bmatrix} 3 \\ 4 \\ -9 \end{bmatrix}$ do exemplo 4.31 também está ilustrado na figura 4.2. Vimos, naquele exemplo, que $\mathbf{v} = 5\mathbf{a}_1 - 2\mathbf{a}_2$ (verifique, caso não o tenha feito ainda). Isso quer dizer que as coordenadas de $\mathbf{v}$ com relação à base $\mathcal{B}$ são 5 e -2, isto é, $[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} 5 \\ -2 \end{bmatrix}$. É interessante verificar a igualdade $\mathbf{v} = 5\mathbf{a}_1 - 2\mathbf{a}_2$ geometricamente, aplicando a regra do paralelogramo sobre figura 4.2.

Similarmente, no exemplo 4.16, vimos que $\mathbf{a}_3 = -4\mathbf{a}_1 + 3\mathbf{a}_2$. (Verifique isso geometricamente, usando, mais uma vez, a figura 4.2 e a regra do paralelogramo.) Assim, o vetor de coordenadas de $\mathbf{a}_3$ com relação à base $\mathcal{B}$ é $[\mathbf{a}_3]_{\mathcal{B}} = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$.

Observe, no exemplo acima, que $\mathbf{v}$ é um vetor de $\mathbb{R}^3$ (assim como $\mathbf{a}_3$), mas que $[\mathbf{v}]_{\mathcal{B}}$ é um vetor de apenas *duas* coordenadas (assim como $[\mathbf{a}_3]_{\mathcal{B}}$). Por estranha que essa situação possa parecer, ela é consistente com a definição 4.32, já que a base $\mathcal{B}$ do plano W possui apenas *dois* elementos, isto é, W tem dimensão igual a 2 (veja também a observação 4.33). Intuitivamente, o que ocorre é que, já que o vetor $\mathbf{v}$ está contido no plano W, podemos especificá-lo usando um sistema de coordenadas *sobre esse plano*. Não é necessário usar um sistema de coordenadas no espaço $\mathbb{R}^3$ inteiro para especificar vetores contidos em W. Por outro lado, um sistema de coordenadas sobre o plano W não é suficiente para especificar vetores quaisquer de $\mathbb{R}^3$.

A situação descrita acima tem uma analogia com um exemplo mais concreto. Vivemos num universo de *três* dimensões espaciais (semelhante ao $\mathbb{R}^3$). No entanto, para especificar um ponto sobre a superfície do planeta Terra, bastam *duas* coordenadas: sua *latitude* e sua *longitude*. Os paralelos e meridianos definem um sistema de coordenadas sobre a superfície da Terra, assim como a base $\{\mathbf{a}_1, \mathbf{a}_2\}$ define um sistema de coordenadas sobre o plano W. Essa analogia, no entanto, não é perfeita. De fato, ela tem uma falha gravíssima: A superfície da Terra *não é* um subespaço, e o sistema de coordenadas geográficas não é dado por uma "base", no sentido estrito da definição 4.10.

Observação

É interessante notar que definição 4.32 só faz sentido graças ao teorema 4.30. De fato, esse teorema garante a existência e a unicidade do vetor de coordenadas $[\mathbf{v}]_{\mathcal{B}}$, para qualquer $\mathbf{v} \in V$, quando $\mathcal{B}$ é uma base de V. Repare, em particular, que se existisem duas (ou mais) maneiras de escrever $\mathbf{v}$ como uma combinação linear dos elementos de $\mathcal{B}$, a definição de $[\mathbf{v}]_{\mathcal{B}}$ seria ambígua. O teorema 4.30 garante que isso nunca ocorre.

4.5 Bases para Nuc A, Col A e Span$\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$

Os resultados das seções 4.3 e 4.4 atestam a importância do conceito de *base*. Na seção 4.2, vimos que qualquer subespaço de $\mathbb{R}^n$, de fato, possui uma base (teorema 4.20). Não estudamos, no entanto, como construir, "na prática", uma base para um dado subespaço. Esse é o objetivo desta seção. Ilustraremos, por meio de exemplos, como determinar bases para subespaços especificados de diversas formas. Discutiremos também a dimensão desses subespaços.

4.5.1 Uma base para Nuc A

Recorde os exemplos 3.12 e 3.13 da seção 3.2, bem como o exercício P3.10. Vimos que, dada uma matriz A, podemos escrever $\operatorname{Nuc} A$ explicitamente como o *span* de certos vetores. Esses vetores formarão, portanto, um *conjunto gerador* de $\operatorname{Nuc} A$. O método proposto nesses exemplos sempre produzirá, de fato, uma *base* do núcleo, como argumentaremos nesta subseção.

Exemplo 4.38

Vamos obter um conjunto gerador do núcleo da matriz A de (1.13) (página 13):

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

Temos que resolver o sistema $A\mathbf{x} = \mathbf{0}$, a fim de achar uma descrição explícita de $\operatorname{Nuc} A$, seguindo o procedimento da seção 3.2. Felizmente, já obtivemos a forma escalonada reduzida de A: é a matriz B dada em (1.15) (página 18). Assim, o sistema $A\mathbf{x} = \mathbf{0}$ é equivalente a $B\mathbf{x} = \mathbf{0}$[9], cuja matriz completa é

$$\left[B \mid \mathbf{0}\right] = \left[\begin{array}{cccrr|c} \boxed{1} & 0 & -3 & 0 & 5 & 0 \\ 0 & \boxed{1} & 2 & 0 & -3 & 0 \\ 0 & 0 & 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

As variáveis básicas são x_1, x_2 e x_4, e as livres são x_3 e x_5. A descrição vetorial paramétrica do conjunto-solução de $A\mathbf{x} = \mathbf{0}$ é, então, dada por

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 3x_3 - 5x_5 \\ -2x_3 + 3x_5 \\ x_3 \\ 0 \\ x_5 \end{bmatrix} = x_3 \underbrace{\begin{bmatrix} 3 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}}_{\mathbf{u}_1} + x_5 \underbrace{\begin{bmatrix} -5 \\ 3 \\ 0 \\ 0 \\ 1 \end{bmatrix}}_{\mathbf{u}_2} = x_3\mathbf{u}_1 + x_5\mathbf{u}_2. \tag{4.7}$$

Isso mostra que $\{\mathbf{u}_1, \mathbf{u}_2\}$ é um conjunto gerador de $\operatorname{Nuc} A$, onde $\mathbf{u}_1$ e $\mathbf{u}_2$ são os vetores de $\mathbb{R}^5$ indicados acima. Em símbolos, $\operatorname{Nuc} A = \operatorname{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$.

Observação 4.39
Os vetores $\mathbf{u}_1$ e $\mathbf{u}_2$ obtidos no exemplo acima são linearmente independentes, pois teremos $x_3\mathbf{u}_1 + x_5\mathbf{u}_2 = \mathbf{0}$ somente se os pesos x_3 e x_5 forem ambos iguais a zero. A forma mais fácil de perceber isso é examinar as componentes 3 e 5 do vetor $\mathbf{x} = x_3\mathbf{u}_1 + x_5\mathbf{u}_2$, que são x_3 e x_5, respectivamente. Evidentemente, para que $\mathbf{x}$ seja o vetor zero, é necessário que essas componentes sejam nulas.

Assim, o conjunto gerador $\{\mathbf{u}_1, \mathbf{u}_2\}$ é linearmente independente e é, portanto, uma *base* de $\operatorname{Nuc} A$.

O método utilizado no exemplo acima para determinar um conjunto gerador do núcleo de uma matriz sempre produzirá uma base, pois um argumento análogo ao da observação 4.39 será válido em todos os exemplos dessa natureza.

Exemplo 4.40
Determine uma base para o núcleo da matriz

$$C = \begin{bmatrix} 2 & 6 & 2 & -4 & 4 \\ 1 & 3 & -3 & 14 & -10 \\ -1 & -3 & 0 & -2 & 1 \end{bmatrix}.$$

Solução: Mais uma vez, temos que resolver o sistema homogêneo $C\mathbf{x} = \mathbf{0}$. Verifique que a forma escalonada *reduzida* da matriz completa $\left[C \mid \mathbf{0}\right]$ é dada por

$$\left[\begin{array}{ccccr|c} \boxed{1} & 3 & 0 & 2 & -1 & 0 \\ 0 & 0 & \boxed{1} & -4 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

[9] Caso tenha dúvidas quanto a isso, veja o exercício P3.5(b).

Assim, o conjunto-solução do sistema $C\mathbf{x} = \mathbf{0}$ é descrito, em termos das variáveis livres x_2, x_4 e x_5, por

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -3x_2 - 2x_4 + x_5 \\ x_2 \\ 4x_4 - 3x_5 \\ x_4 \\ x_5 \end{bmatrix} = x_2 \underbrace{\begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{\mathbf{v}_1} + x_4 \underbrace{\begin{bmatrix} -2 \\ 0 \\ 4 \\ 1 \\ 0 \end{bmatrix}}_{\mathbf{v}_2} + x_5 \underbrace{\begin{bmatrix} 1 \\ 0 \\ -3 \\ 0 \\ 1 \end{bmatrix}}_{\mathbf{v}_3}$$

$$= x_2\mathbf{v}_1 + x_4\mathbf{v}_2 + x_5\mathbf{v}_3. \quad (4.8)$$

Sejam $\mathbf{v}_1$, $\mathbf{v}_2$ e $\mathbf{v}_3$ os vetores de $\mathbb{R}^5$ indicados acima. O conjunto $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ é uma base de $\operatorname{Nuc} C$: esse conjunto não apenas gera $\operatorname{Nuc} C$, como também é linearmente independente. Como mencionamos, um argumento análogo ao da observação 4.39 justifica a independência linear. Os detalhes são deixados para o exercício P4.21. □

Vamos, agora, discutir a dimensão do núcleo de uma matriz. Primeiro, consideramos os dois exemplos anteriores. A dimensão do núcleo da matriz A do exemplo 4.38 é igual a 2, pois a base $\{\mathbf{u}_1, \mathbf{u}_2\}$ de $\operatorname{Nuc} A$ tem *dois* elementos (assim, pelo teorema 4.23, qualquer base terá dois elementos). Observe que o sistema $A\mathbf{x} = \mathbf{0}$ tem *duas* variáveis livres, x_3 e x_5, que são os coeficientes dos vetores $\mathbf{u}_1$ e $\mathbf{u}_2$ na descrição paramétrica (4.7).

O núcleo da matriz C do exemplo 4.40, por sua vez, tem dimensão igual a 3, pois a base $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ tem *três* elementos. Observe, mais uma vez, que o sistema $C\mathbf{x} = \mathbf{0}$ tem *três* variáveis livres, x_2, x_4 e x_5, que são os coeficientes dos vetores $\mathbf{v}_1$, $\mathbf{v}_2$ e $\mathbf{v}_3$ na descrição (4.8).

Perceba que, para qualquer matriz A, sempre haverá uma correspondência biunívoca entre as variáveis livres do sistema $A\mathbf{x} = \mathbf{0}$ e os vetores da base de $\operatorname{Nuc} A$ construída pelo método dos exemplos acima. Com efeito, cada variável livre será o coeficiente de um dos vetores da base, na descrição vetorial paramétrica do conjunto-solução de $A\mathbf{x} = \mathbf{0}$. Assim, *a dimensão de* $\operatorname{Nuc} A$ *será igual ao número de variáveis livres do sistema* $A\mathbf{x} = \mathbf{0}$.

Lembre-se de que as variáveis livres do sistema $A\mathbf{x} = \mathbf{0}$ são aquelas que correspondem às *colunas não-pivô* da matriz de coeficientes A. Assim, vale o seguinte resultado.

Proposição 4.41
Seja A uma matriz qualquer. A dimensão do núcleo de A é igual ao número de colunas não-pivô de A.

A dimensão do núcleo de A, que denotamos por $\dim \operatorname{Nuc} A$, é às vezes chamada de **nulidade** da matriz A.

4.5.2 Uma base para Col A

Veremos, nesta subseção, como obter uma base para o espaço das colunas de uma matriz. Comecemos com um exemplo simples.

Exemplo 4.42
Vamos encontrar uma base para $\operatorname{Col} B$, onde B é a matriz de (1.15) (página 18):

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

Denotamos as colunas de B por $\mathbf{b}_1$, $\mathbf{b}_2$, ... , $\mathbf{b}_5$. A matriz B está na forma escalonada reduzida, então é fácil ver que as colunas não-pivô $\mathbf{b}_3$ e $\mathbf{b}_5$ são combinações lineares das demais. De fato, podemos "ler", diretamente da matriz B, as relações $\mathbf{b}_3 = -3\mathbf{b}_1 + 2\mathbf{b}_2$ e $\mathbf{b}_5 = 5\mathbf{b}_1 - 3\mathbf{b}_2$.

Por definição, $\{\mathbf{b}_1, \dots, \mathbf{b}_5\}$ é um conjunto gerador do espaço-coluna de B, isto é, $\operatorname{Col} B = \operatorname{Span}\{\mathbf{b}_1, \dots, \mathbf{b}_5\}$. Em vista do lema 4.17, podemos remover os vetores "redundantes" $\mathbf{b}_3$ e $\mathbf{b}_5$, e o conjunto resultante $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_4\}$ ainda irá gerar $\operatorname{Col} B$. Se quiser, verifique esse fato diretamente, usando um argumento similar ao do exemplo 4.16. Verifique, também, que o conjunto $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_4\}$ é linearmente independente (veja o exercício P4.23). Isso mostra que esse conjunto é uma base de $\operatorname{Col} B$.

O exemplo acima é particularmente simples porque a matriz B está na forma escalonada reduzida. Antes de passar a um exemplo mais interessante, precisamos do resultado a seguir.

Proposição 4.43
Sejam A e B matrizes linha-equivalentes. Então os vetores-coluna de A e os vetores-coluna de B possuem as mesmas relações de dependência linear.

Demonstração: Se A e B são linha-equivalentes, então as matrizes $[A \mid \mathbf{0}]$ e $[B \mid \mathbf{0}]$ também são linha-equivalentes (veja o exercício P3.5(b)). Assim, os sistemas $A\mathbf{x} = \mathbf{0}$ e $B\mathbf{x} = \mathbf{0}$ têm o mesmo conjunto-solução. Para terminar a demonstração, basta lembrar que as soluções não-triviais de um sistema $A\mathbf{x} = \mathbf{0}$ correspondem precisamente às relações de dependência linear entre as colunas de A (veja o exercício P3.18). $\square$

O significado dessa proposição ficará mais claro no exemplo abaixo.

Exemplo 4.44
Vamos achar uma base para $\operatorname{Col} A$, onde A é a matriz de (1.13) (página 13):

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

Denotamos as colunas de A por $\mathbf{a}_1$, $\mathbf{a}_2$, ... , $\mathbf{a}_5$. Observe que a matriz A é linha-equivalente à matriz B do exemplo anterior (de fato, B foi obtida de A por operações-linha, na seção 1.5). Pela proposição 4.43, as colunas $\mathbf{a}_1$, ... , $\mathbf{a}_5$ da matriz A têm as mesmas relações de dependência linear que as colunas $\mathbf{b}_1, \dots, \mathbf{b}_5$ da matriz B. Já havíamos observado que $\mathbf{b}_3 = -3\mathbf{b}_1 + 2\mathbf{b}_2$ e que $\mathbf{b}_5 = 5\mathbf{b}_1 - 3\mathbf{b}_2$. Assim, valem também as relações $\mathbf{a}_3 = -3\mathbf{a}_1 + 2\mathbf{a}_2$ e $\mathbf{a}_5 = 5\mathbf{a}_1 - 3\mathbf{a}_2$ (verifique-as diretamente!).

Dessa maneira, podemos remover os vetores "redundantes" $\mathbf{a}_3$ e $\mathbf{a}_5$ do conjunto $\{\mathbf{a}_1, \dots, \mathbf{a}_5\}$, e o conjunto resultante $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_4\}$ ainda irá gerar $\operatorname{Col} A$. Esse conjunto é também linearmente independente, pelo seguinte argumento. Se houvesse uma relação de dependência linear entre os vetores $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{a}_4$, essa mesma relação valeria para os vetores $\mathbf{b}_1$, $\mathbf{b}_2$ e $\mathbf{b}_4$, pela proposição 4.43, novamente. Mas já constatamos que $\mathbf{b}_1$, $\mathbf{b}_2$ e $\mathbf{b}_4$ são linearmente independentes.

Assim, mostramos que o conjunto $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_4\}$ é uma base de $\operatorname{Col} A$. Observe que esse é o conjunto formado pelas *colunas-pivô* da matriz A.

Os exercícios P4.22 e P4.24 contêm um roteiro para generalizar os exemplos acima, de forma a demonstrar o resultado a seguir.

Proposição 4.45
As colunas-pivô de uma matriz constituem uma base de seu espaço-coluna.

Para achar uma base para o espaço-coluna de uma matriz, portanto, basta "desvendar", via escalonamento, quais são suas colunas-pivô.

Exemplo 4.46
Encontre uma base para o espaço das colunas da matriz

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

Solução: Vamos escalonar a matriz C:

$$\begin{bmatrix} \boxed{1} & -2 & 8 \\ 0 & 2 & -4 \\ 3 & 2 & 8 \end{bmatrix} \xrightarrow{\ell_3 \to \ell_3 - 3\ell_1} \begin{bmatrix} \boxed{1} & -2 & 8 \\ 0 & \boxed{2} & -4 \\ 0 & 8 & -16 \end{bmatrix} \xrightarrow{\ell_3 \to \ell_3 - 4\ell_2} \begin{bmatrix} \boxed{1} & -2 & 8 \\ 0 & \boxed{2} & -4 \\ 0 & 0 & 0 \end{bmatrix}. \tag{4.9}$$

Observe que não é necessário obter a forma escalonada reduzida. Constatamos que as colunas-pivô da matriz C são a primeira e a segunda. Assim, pela proposição 4.45, uma base de $\operatorname{Col} C$ é dada pelo conjunto

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} -2 \\ 2 \\ 2 \end{bmatrix} \right\}. \tag{4.10}$$

Verifique que, de fato, a terceira coluna de C é gerada pelas duas primeiras. □

Cuidado!
Na busca de uma base para o espaço-coluna de uma dada matriz, o escalonamento serve apenas para determinar *quais são* as suas colunas-pivô. Ao escrever a base, tenha o cuidado de usar as colunas *da própria matriz dada*, e não as da forma

escalonada obtida. Observe, no exemplo 4.46, que a base (4.10) de $\operatorname{Col} C$ é composta pelas colunas-pivô *da matriz* C, e não pelas colunas-pivô $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ e $\begin{bmatrix} -2 \\ 2 \\ 0 \end{bmatrix}$ da matriz escalonada em (4.9). Estes dois vetores, de fato, *não geram* o espaço-coluna de C. Note, em particular, que eles têm zero como último elemento, e, portanto, não podem gerar nenhum dos vetores-coluna de C. Similarmente, a base de $\operatorname{Col} A$ obtida no exemplo 4.44 contém as colunas $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{a}_4$ *da própria matriz* A, e não as colunas $\mathbf{b}_1$, $\mathbf{b}_2$ e $\mathbf{b}_4$ da forma escalonada B.

Definição 4.47
Seja A uma matriz qualquer. A dimensão do espaço das colunas de A é chamado de **posto** de A, e é denotado por $\operatorname{posto} A$. Ou seja, definimos $\operatorname{posto} A = \dim \operatorname{Col} A$.

Exemplo 4.48
O posto da matriz A do exemplo 4.44 é igual a 3, visto que $\operatorname{Col} A$ tem uma base com três vetores. Observe que A tem três colunas-pivô, que formam, justamente, a base de $\operatorname{Col} A$ obtida no exemplo 4.44. Já o posto da matriz C do exemplo 4.46 é igual a 2, visto que $\operatorname{Col} C$ tem uma base com dois vetores (as colunas-pivô de C). Em símbolos, $\operatorname{posto} A = 3$ e $\operatorname{posto} C = 2$.

Como vimos, as colunas-pivô de qualquer matriz A formam uma base para seu espaço-coluna. Assim, a dimensão de $\operatorname{Col} A$ é sempre igual ao número de colunas-pivô de A. Em outras palavras, temos o resultado abaixo.

Proposição 4.49
O posto de uma matriz A é igual número de colunas-pivô de A.

Alguns autores denotam o posto de A por $\operatorname{rank} A$, pois, em inglês, a palavra usada para *posto* é *rank*.

4.5.3 Uma base para $\operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$

No exemplo 4.16 (página 89), verificamos que o conjunto $\mathcal{B} = \{\mathbf{a}_1, \mathbf{a}_2\}$ é uma base do subespaço $W \subseteq \mathbb{R}^3$ gerado pelos vetores $\mathbf{a}_1$, $\mathbf{a}_2$ e $\mathbf{a}_3$ dados. No entanto, não discutimos ainda o caso geral. Dada uma lista $\mathbf{a}_1$, $\mathbf{a}_2$, ... , $\mathbf{a}_m$ de vetores de $\mathbb{R}^n$, como podemos obter, de maneira sistemática, uma base para o subespaço $\operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ por eles gerado?

É fácil abordar essa questão usando os resultados da subseção anterior. Primeiro, consideramos a matriz $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_m \end{bmatrix}$, cujas colunas são os vetores $\mathbf{a}_j$ dados. Dessa maneira, teremos $\operatorname{Col} A = \operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$, pela definição de espaço-coluna. Agora, basta aplicar a proposição 4.45: as colunas-pivô de A formarão uma base de $\operatorname{Col} A$ e, portanto, também de $\operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$, já que esses subespaços são iguais.

Exemplo 4.50
Encontre uma base para o subespaço de $\mathbb{R}^4$ gerado pelos vetores

$$\mathbf{v}_1 = \begin{bmatrix} -1 \\ 3 \\ 2 \\ 4 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ -4 \\ 3 \\ 3 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -2 \\ 5 \\ 9 \\ 15 \end{bmatrix} \quad \text{e} \quad \mathbf{v}_4 = \begin{bmatrix} 1 \\ 2 \\ 2 \\ -1 \end{bmatrix}.$$

Em seguida, determine a dimensão desse subespaço.

Solução: Seja $A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 \end{bmatrix}$, de forma que $\operatorname{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_4\} = \operatorname{Col} A$. Vamos escalonar a matriz A para revelar suas colunas-pivô:

$$A = \begin{bmatrix} \boxed{-1} & 1 & -2 & 1 \\ 3 & -4 & 5 & 2 \\ 2 & 3 & 9 & 2 \\ 4 & 3 & 15 & -1 \end{bmatrix} \longrightarrow \begin{bmatrix} \boxed{-1} & 1 & -2 & 1 \\ 0 & \boxed{-1} & -1 & 5 \\ 0 & 5 & 5 & 4 \\ 0 & 7 & 7 & 3 \end{bmatrix} \longrightarrow$$

$$\longrightarrow \begin{bmatrix} \boxed{-1} & 1 & -2 & 1 \\ 0 & \boxed{-1} & -1 & 5 \\ 0 & 0 & 0 & \boxed{29} \\ 0 & 0 & 0 & 38 \end{bmatrix} \longrightarrow \begin{bmatrix} \boxed{-1} & 1 & -2 & 1 \\ 0 & \boxed{-1} & -1 & 5 \\ 0 & 0 & 0 & \boxed{29} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Não é necessário chegar à forma reduzida. As colunas-pivô de A são a primeira, a segunda e a quarta. Dessa forma, o conjunto $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4\}$ é uma base do subespaço $\operatorname{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_4\}$. Como a base encontrada tem três vetores, a dimensão desse subespaço é igual a 3. □

4.6 O teorema do posto

Como vimos na proposição 4.49, o posto de uma matriz A (a dimensão de seu espaço-coluna) é igual ao número de colunas-pivô de A. Por outro lado, a proposição 4.41 diz que a dimensão do núcleo de A é dada pelo número de colunas não-pivô de A, ou seja, pelo número de variáveis livres do sistema $A\mathbf{x} = \mathbf{0}$. Em resumo:

$$\begin{aligned} \operatorname{posto} A &= \text{número de colunas-pivô de } A, \\ \dim \operatorname{Nuc} A &= \text{número de colunas não-pivô de } A. \end{aligned}$$

Ora, a soma do número de colunas-pivô e não-pivô de A resulta no número total de colunas de A! Esta observação trivial leva ao resultado abaixo.

Teorema 4.51 (Teorema do Posto, versão 1)
Se A é uma matriz com m colunas, então

$$\operatorname{posto} A + \dim \operatorname{Nuc} A = m. \tag{4.11}$$

Enfatizamos que (4.11) é nada mais do que uma reformulação do fato trivial

$$\begin{pmatrix} \text{número de} \\ \text{colunas-pivô de } A \end{pmatrix} + \begin{pmatrix} \text{número de colunas} \\ \text{não-pivô de } A \end{pmatrix} = \begin{pmatrix} \text{número total de} \\ \text{colunas de } A \end{pmatrix}.$$

Esse fato, no entanto, tem aplicações interessantes, e é conveniente ter a forma sintética (4.11) de exprimi-lo. O teorema do posto será revisitado em um contexto um pouco diferente na subseção 5.3.3, por isso usamos aqui o rótulo "versão 1".

4.7 Demonstrações dos resultados sobre bases e dimensão (leitura opcional)

A seguir, fornecemos as demonstrações que omitimos nas seções 4.2 e 4.3.

Demonstração do lema 4.17: Evidentemente, os elementos de $\mathcal{G}'$ pertencem a V, visto que $\mathcal{G}'$ está contido em $\mathcal{G}$, e este é um conjunto gerador de V (portanto está, ele próprio, contido em V). Assim, a condição (i) da observação 4.9 está satisfeita. Basta mostrar, então, que qualquer vetor $\mathbf{v}$ de V é gerado pelos vetores de $\mathcal{G}'$. O vetor $\mathbf{v}$ pode ser escrito na forma

$$\mathbf{v} = c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_{k-1}\mathbf{a}_{k-1} + c_k\mathbf{a}_k + c_{k+1}\mathbf{a}_{k+1} + \cdots + c_m\mathbf{a}_m, \tag{4.12}$$

já que $\mathbf{v} \in V$ e $\mathcal{G}$ gera V. Agora, por hipótese, $\mathbf{a}_k$ é uma combinação linear dos demais vetores de $\mathcal{G}$. Assim, o vetor $\mathbf{a}_k$ pode ser escrito como

$$\mathbf{a}_k = d_1\mathbf{a}_1 + d_2\mathbf{a}_2 + \cdots + d_{k-1}\mathbf{a}_{k-1} + d_{k+1}\mathbf{a}_{k+1} + \cdots + d_m\mathbf{a}_m. \tag{4.13}$$

Substituindo a expressão (4.13) para $\mathbf{a}_k$ em (4.12) e simplificando, obtemos

$$\begin{aligned}\mathbf{v} = (c_1 + c_kd_1)\mathbf{a}_1 + (c_2 + c_kd_2)\mathbf{a}_2 + \cdots + (c_{k-1} + c_kd_{k-1})\mathbf{a}_{k-1} +&\\ + (c_{k+1} + c_kd_{k+1})\mathbf{a}_{k+1} + \cdots + (c_m + c_kd_m)\mathbf{a}_m.&\end{aligned}$$

Mostramos, então, que $\mathbf{v}$ é gerado pelos vetores de $\mathcal{G}'$. □

Demonstração do teorema 4.18: Se $\mathcal{G}$ é um conjunto linearmente independente, então ele já é uma base de V, pois, por hipótese, ele é também um conjunto gerador. Não há mais o que considerar nesse caso.

Se, do contrário, $\mathcal{G}$ é linearmente *dependente*, então, pela proposição 3.18, um dos vetores de $\mathcal{G}$ é uma combinação linear dos demais. Pelo lema 4.17, podemos retirar esse vetor “redundante” do conjunto $\mathcal{G}$, e o conjunto $\mathcal{G}'$ obtido ainda irá gerar V. Se $\mathcal{G}'$ for linearmente independente, ele será uma base de V, e o processo terminará. Se, por outro lado, $\mathcal{G}'$ for linearmente dependente, poderemos mais uma vez retirar um vetor de $\mathcal{G}'$, de forma que o conjunto resultante ainda gere V. O processo continuará até que obtenhamos um conjunto linearmente independente. Como temos conjuntos geradores em cada etapa, o conjunto obtido irá também gerar V, ou seja, será uma base de V.

Para terminar a prova de forma honesta, no entanto, precisamos justificar por que esse processo, de fato, termina. Por que o processo não poderia continuar para sempre, sem sucesso? Como podemos garantir que, em algum estágio, vamos necessariamente obter um conjunto linearmente independente? Ora, o conjunto $\mathcal{G}$ dado no enunciado tem um número finito m de elementos. Assim, o processo descrito acima poderá ter, no máximo, m etapas de remoção de vetores. Se todos os m vetores de $\mathcal{G}$ forem realmente removidos pelo processo, terminaremos com o conjunto vazio, que é linearmente independente, conforme a observação 4.12. (Este caso extremo só ocorre quando V é o subespaço trivial.) □

Demonstração do teorema 4.19: Vamos fazer algumas considerações preliminares. O conjunto $\mathcal{I}$ é linearmente independente, e seus elementos são vetores de $\mathbb{R}^n$, já que $\mathcal{I} \subseteq V \subseteq \mathbb{R}^n$. Pela proposição 3.22, $\mathcal{I}$ contém no máximo n vetores. Assim, $\mathcal{I}$ pode ser escrito como $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$, com $m \leqslant n$, ou então como $\{\}$, no caso em que $\mathcal{I}$ é o conjunto vazio. Por hipótese, os vetores de $\mathcal{I}$ pertencem a V, como já indicamos. Pela observação 4.2, portanto, vale

$$\operatorname{Span} \mathcal{I} \subseteq V, \tag{4.14}$$

onde $\operatorname{Span} \mathcal{I}$ representa o conjunto gerado pelos vetores de $\mathcal{I}$. Agora vamos passar ao cerne da prova do teorema.

Se $\mathcal{I}$ é um conjunto gerador de V, então ele já é uma base de V (pois é também linearmente independente). Neste caso, a relação (4.14) vale com igualdade ($\operatorname{Span} \mathcal{I} = V$), e não há mais trabalho algum a fazer.

Vamos considerar, então, o caso em que $\mathcal{I} = \{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ *não* é um conjunto gerador de V, ou seja, o caso em que a inclusão (4.14) é *estrita* ($\operatorname{Span} \mathcal{I} \subsetneq V$).[10] Nesse caso, podemos escolher um vetor $\mathbf{v}$ de V que *não pertença* a $\operatorname{Span} \mathcal{I}$. Pelo exercício resolvido R3.3, o conjunto $\mathcal{I}' = \{\mathbf{a}_1, \ldots, \mathbf{a}_m, \mathbf{v}\}$, obtido pela inclusão do vetor $\mathbf{v}$ em $\mathcal{I}$, ainda será linearmente independente. Observe que $\operatorname{Span} \mathcal{I}' \subseteq V$, pois todos os vetores de $\mathcal{I}'$ pertencem a V. Se o conjunto $\mathcal{I}'$ gerar V, isto é, se tivermos a igualdade $\operatorname{Span} \mathcal{I}' = V$, esse conjunto será uma base de V, e processo terminará. Caso contrário, poderemos acrescentar a $\mathcal{I}'$ mais um vetor de V, de forma que o conjunto resultante ainda seja linearmente independente. O processo continuará até que obtenhamos um conjunto gerador de V. Como temos conjuntos linearmente independentes em cada etapa, o conjunto gerador obtido será uma base de V.

Mais uma vez, é preciso justificar por que o processo necessariamente termina. Como é que podemos garantir que, em alguma etapa, um conjunto gerador para V será obtido? Ora, se um conjunto gerador nunca fosse obtido, poderíamos continuar acrescentando a $\mathcal{I}$ um número arbitrário de vetores, um por um, *sempre mantendo a propriedade de independência linear dos conjuntos resultantes.* Mas isso é impossível, em razão da proposição 3.22. $\square$

Demonstração do lema 4.21: Suponha que $\mathcal{G} = \{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ seja um conjunto gerador de V. Seja $\mathcal{S} = \{\mathbf{v}_1, \ldots, \mathbf{v}_q\}$ qualquer conjunto de vetores de V, com $q > m$. Vamos provar que $\mathcal{S}$ tem que ser linearmente dependente.

Como $\mathcal{G}$ gera V, cada vetor $\mathbf{v}_i$ de $\mathcal{S}$ pode ser escrito como uma combinação linear dos vetores $\mathbf{a}_j$ de $\mathcal{G}$.[11] Sendo assim, existem escalares u_{ij} tais que

$$\begin{aligned} \mathbf{v}_1 &= u_{11}\mathbf{a}_1 + u_{12}\mathbf{a}_2 + \cdots + u_{1m}\mathbf{a}_m, \\ \mathbf{v}_2 &= u_{21}\mathbf{a}_1 + u_{22}\mathbf{a}_2 + \cdots + u_{2m}\mathbf{a}_m, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \mathbf{v}_q &= u_{q1}\mathbf{a}_1 + u_{q2}\mathbf{a}_2 + \cdots + u_{qm}\mathbf{a}_m. \end{aligned} \tag{4.15}$$

[10] Em outras palavras, $\operatorname{Span} \mathcal{I}$ está *propriamente contido* em V (não vale a igualdade).

[11] A hipótese de que os vetores de $\mathcal{S}$ pertençam a V é essencial nesse passo!

Vamos agora definir os seguintes q vetores de $\mathbb{R}^m$:

$$\mathbf{u}_1 = \begin{bmatrix} u_{11} \\ u_{12} \\ \vdots \\ u_{1m} \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} u_{21} \\ u_{22} \\ \vdots \\ u_{2m} \end{bmatrix}, \quad \dots, \quad \mathbf{u}_q = \begin{bmatrix} u_{q1} \\ u_{q2} \\ \vdots \\ u_{qm} \end{bmatrix}.$$

Não há nenhum mistério aqui. Apenas atribuímos os nomes $\mathbf{u}_1$, $\mathbf{u}_2$, ... , $\mathbf{u}_q$ aos vetores indicados acima. Vamos também chamar de A a matriz cujas colunas são os vetores $\mathbf{a}_j$, isto é, $A = \begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_m\end{bmatrix}$. Com esta notação, as q equações de (4.15) podem ser reescritas como

$$\mathbf{v}_1 = A\mathbf{u}_1, \qquad \mathbf{v}_2 = A\mathbf{u}_2, \qquad \dots, \qquad \mathbf{v}_q = A\mathbf{u}_q. \tag{4.16}$$

Para verificar a equivalência entre (4.15) e (4.16), basta usar a definição 2.4 do produto matriz-vetor (página 43).

Observe que $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_q\}$ é um conjunto de q vetores de $\mathbb{R}^m$ e que, por hipótese, $q > m$. Pela proposição 3.22, esse conjunto é linearmente dependente. Dessa forma, existem escalares $c_1, \dots, c_q$, *não todos iguais a zero*, tais que $c_1\mathbf{u}_1 + \cdots + c_q\mathbf{u}_q = \mathbf{0}_m$. Essa relação de dependência linear se traduz para os vetores $\mathbf{v}_i$, por meio das equações (4.16). Com efeito, temos

$$\begin{aligned} c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_q\mathbf{v}_q &= c_1A\mathbf{u}_1 + c_2A\mathbf{u}_2 + \cdots + c_qA\mathbf{u}_q = \\ &= A(c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_q\mathbf{u}_q) = A\mathbf{0}_m = \mathbf{0}_n. \end{aligned}$$

Isso mostra que o conjunto $\mathcal{S} = \{\mathbf{v}_1, \dots, \mathbf{v}_q\}$ é linearmente dependente. □

Demonstração do teorema 4.26: Suponha que $\mathcal{S}$ seja linearmente independente. Se $\mathcal{S}$ não gerasse V, então, pelo teorema 4.19, poderíamos acrescentar vetores de V a $\mathcal{S}$ de forma a obter uma base. Essa base, no entanto, conteria *mais* do que p vetores, pois seria obtida *acrescentando* vetores a $\mathcal{S}$, que já tem p elementos. Isso contradiria a hipótese de que p seja a dimensão de V. Assim, $\mathcal{S}$ tem que ser um conjunto gerador de V. Isso termina a prova de (a).

Agora, suponha que $\mathcal{S}$ gere V. Se $\mathcal{S}$ não fosse linearmente independente, então, pelo teorema 4.18, poderíamos remover elementos de $\mathcal{S}$ de forma a obter uma base de V. Essa base, no entanto, conteria *menos* do que p vetores. Mais uma vez, isso contradiria a hipótese $\dim V = p$. Portanto, $\mathcal{S}$ tem que ser linearmente independente. Isso prova (b). □

Exercícios resolvidos

R4.1. Sejam $\mathbf{b}_1 = \left[\begin{smallmatrix} 2 \\ -1 \end{smallmatrix}\right]$ e $\mathbf{b}_2 = \left[\begin{smallmatrix} -3 \\ 2 \end{smallmatrix}\right]$, e considere a base $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ de $\mathbb{R}^2$. Determine o vetor $\mathbf{x} \in \mathbb{R}^2$ tal que $[\mathbf{x}]_{\mathcal{B}} = \left[\begin{smallmatrix} 4 \\ 2 \end{smallmatrix}\right]$.

Solução: Basta aplicar a definição 4.32. Dizer que $[\mathbf{x}]_{\mathcal{B}} = \left[\begin{smallmatrix} 4 \\ 2 \end{smallmatrix}\right]$ é o mesmo que dizer $\mathbf{x} = 4\mathbf{b}_1 + 2\mathbf{b}_2$, ou seja, $\mathbf{x} = 4\left[\begin{smallmatrix} 2 \\ -1 \end{smallmatrix}\right] + 2\left[\begin{smallmatrix} -3 \\ 2 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 2 \\ 0 \end{smallmatrix}\right]$. Note que 2 e 0 são as coordenadas de $\mathbf{x}$ com respeito à base canônica de $\mathbb{R}^2$. □

R4.2. Mostre que se V e W são subespaços de $\mathbb{R}^n$, então a interseção $V \cap W$ é também um subespaço de $\mathbb{R}^n$.

Solução: Vamos verificar que o subconjunto $V \cap W$ satisfaz as condições da definição 4.1. É claro que $V \cap W$ é um subconjunto de $\mathbb{R}^n$, já que V e W o são. O vetor zero de $\mathbb{R}^n$ pertence a V e a W (justifique), portanto ele pertence à interseção $V \cap W$. Sejam agora $\mathbf{u}$ e $\mathbf{v}$ quaisquer vetores de $V \cap W$. Isso significa que cada um desses vetores pertence a V e a W. Assim, a soma $\mathbf{u} + \mathbf{v}$ pertence a V e a W (justifique novamente), portanto pertence a $V \cap W$. A verificação da última condição é deixada para o leitor. □

R4.3. Seja V um subespaço de $\mathbb{R}^n$. Verifique as seguintes afirmativas:

(a) Se V contiver m vetores linearmente independentes, então $\dim V \geqslant m$.
(b) Se V for gerado por um conjunto contendo q vetores, então $\dim V \leqslant q$.

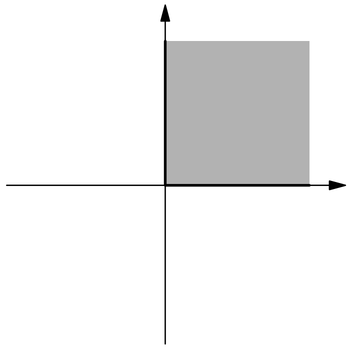
Solução: Este exercício é uma mera reformulação da proposição 4.22. Verifique, lembrando que, se V tem uma base contendo p elementos, então, por definição, $\dim V = p$. □

Observação: Na proposição 4.22, não empregamos explicitamente o termo "dimensão". De fato, essa proposição foi essencial para definir o termo, de modo que empregá-lo seria incoerente do ponto de vista lógico. No entanto, uma vez que o conceito de dimensão esteja adequadamente definido, temos a liberdade de reformular a proposição. O objetivo deste exercício foi exatamente esse.

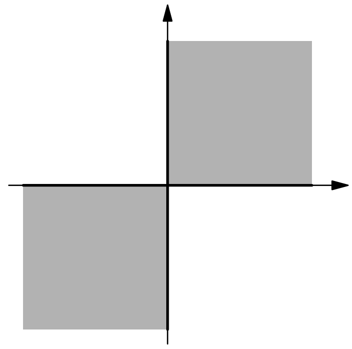
Exercícios propostos

P4.1. Em cada item abaixo, temos um subconjunto do $\mathbb{R}^2$. Suponha que cada um deles contenha a sua respectiva fronteira (representada pela linha mais grossa). Esses subconjuntos *não* são subespaços de $\mathbb{R}^2$. Determine qual(is) propriedade(s) da definição 4.1 é(são) violada(s) em cada caso. Justifique suas respostas graficamente, esboçando algumas figuras.

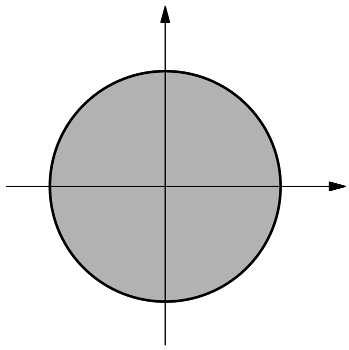
(a) O primeiro quadrante de $\mathbb{R}^2$.



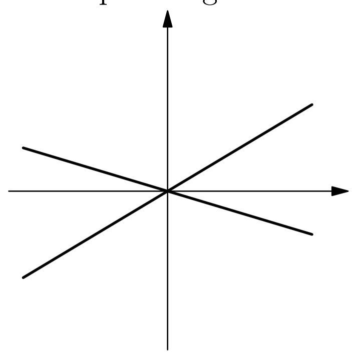
(b) A união do primeiro quadrante com o terceiro quadrante.



(c) Um disco com centro na origem.



(d) A união de duas retas distintas passando pela origem.



P4.2. Mostre que se V é um subespaço de $\mathbb{R}^n$ e os vetores $\mathbf{a}_1, \ldots, \mathbf{a}_m$ pertencem a V, então qualquer combinação linear $c_1\mathbf{a}_1 + \cdots + c_m\mathbf{a}_m$ desses vetores também pertence a V, conforme a observação 4.2.

Dicas: Primeiro, observe que os vetores $c_1\mathbf{a}_1, \ldots, c_m\mathbf{a}_m$ pertencem a V, pela condição 3 da definição 4.1. Em seguida, aplique repetidamente a condição 2.

P4.3. Seja $Q = \{\mathbf{x} \in \mathbb{R}^2;\ x_1 \geqslant 0,\ x_2 \geqslant 0\}$ o primeiro quadrante de $\mathbb{R}^2$, incluindo os semi-eixos positivos (ou seja, Q é o subconjunto dado no item (a) do exercício P4.1). Sejam também $\mathbf{e}_1 = \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ e $\mathbf{e}_2 = \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$. Mostre que $\mathbf{e}_1$ e $\mathbf{e}_2$ pertencem a Q, mas $\operatorname{Span}\{\mathbf{e}_1, \mathbf{e}_2\} = \mathbb{R}^2 \not\subseteq Q$. Explique por que isso *não* contradiz a observação 4.2.

P4.4. Em cada item, determine se os vetores dados formam uma base de $\mathbb{R}^2$. Justifique.

(a) $\begin{bmatrix}3\\1\end{bmatrix}, \begin{bmatrix}-1\\2\end{bmatrix}$.   (b) $\begin{bmatrix}3\\-6\end{bmatrix}, \begin{bmatrix}-1\\2\end{bmatrix}$.   (c) $\begin{bmatrix}3\\1\end{bmatrix}, \begin{bmatrix}-1\\2\end{bmatrix}, \begin{bmatrix}4\\3\end{bmatrix}$.

P4.5. Em cada item, determine se os vetores dados formam uma base de $\mathbb{R}^3$. Justifique.

(a) $\begin{bmatrix} -1 \\ 3 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 \\ 1 \\ -4 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \\ 1 \end{bmatrix}$. (b) $\begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ -8 \\ 5 \end{bmatrix}$. (c) $\begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix}$.

P4.6. Sejam $\mathbf{e}_1, \ldots, \mathbf{e}_n$ os vetores dados em (4.2) (página 86). Verifique, *sem utilizar a independência linear dos vetores* $\mathbf{e}_j$, que (4.3) é a *única* forma de representar um vetor $\mathbf{v}$ de $\mathbb{R}^n$ como combinação linear de $\mathbf{e}_1, \ldots, \mathbf{e}_n$.

Dica: Escreva duas combinações lineares dos vetores $\mathbf{e}_j$ e mostre, diretamente, que elas serão iguais somente se os pesos correspondentes forem todos iguais. Para isso, use a própria equação (4.3).

P4.7. Mostre que os vetores $\mathbf{e}_1, \ldots, \mathbf{e}_n$ de (4.2) são linearmente independentes.

Dica: Use a equação (4.3) e a definição 3.14 de independência linear.

P4.8. Verifique que o lema 4.17 pode ser reformulado da seguinte maneira "compacta": Se $\mathbf{a}_k \in \operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_{k-1}, \mathbf{a}_{k+1}, \ldots, \mathbf{a}_m\}$, então vale a igualdade $\operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_{k-1}, \mathbf{a}_{k+1}, \ldots, \mathbf{a}_m\} = \operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$. (Repare que essa formulação não faz referência alguma a subespaços "abstratos".)

P4.9. O teorema 4.18 diz que uma base de um subespaço pode ser obtida mediante a remoção de elementos "redundantes" de um conjunto gerador (releia o enunciado preciso, na página 90). Este exercício mostra que, em geral, há uma certa liberdade na escolha do(s) elemento(s) a ser(em) removido(s), mas isso não é verdade em todos os casos. Considere, primeiro, os vetores $\mathbf{a}_1 = \left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$, $\mathbf{a}_2 = \left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right]$ e $\mathbf{a}_3 = \left[\begin{smallmatrix} -1 \\ 0 \end{smallmatrix}\right]$ de $\mathbb{R}^2$ e o conjunto $\mathcal{G} = \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$.

(a) Verifique que $\mathcal{G}$ é um conjunto gerador de $\mathbb{R}^2$, mas não é uma base.
(b) Mostre que uma base de $\mathbb{R}^2$ pode ser obtida de $\mathcal{G}$ mediante a remoção de *qualquer um* dos vetores $\mathbf{a}_j$. Em outras palavras, verifique que os conjuntos $\{\mathbf{a}_1, \mathbf{a}_2\}$, $\{\mathbf{a}_1, \mathbf{a}_3\}$ e $\{\mathbf{a}_2, \mathbf{a}_3\}$ são bases de $\mathbb{R}^2$.
(c) Conclua que nenhum dos vetores $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ é "intrinsecamente redundante" no conjunto gerador $\mathcal{G}$, ou seja, nenhum deles é "mais redundante" do que os outros.

Agora, considere o vetor $\mathbf{a}_4 = \left[\begin{smallmatrix} 2 \\ 4 \end{smallmatrix}\right]$ e o conjunto $\mathcal{G}' = \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_4\}$.

(d) Verifique que $\mathcal{G}'$ gera $\mathbb{R}^2$, mas não é uma base.
(e) Mostre que uma base de $\mathbb{R}^2$ pode ser obtida de $\mathcal{G}'$ mediante a remoção de $\mathbf{a}_1$ ou de $\mathbf{a}_4$, *mas não de* $\mathbf{a}_2$.

Finalmente, considere o vetor $\mathbf{a}_5 = \mathbf{0}_2$ e o conjunto $\mathcal{G}'' = \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_5\}$.

(f) Verifique que $\mathcal{G}''$ gera $\mathbb{R}^2$, mas não é uma base.
(g) Mostre que uma base de $\mathbb{R}^2$ pode ser obtida de $\mathcal{G}''$ mediante a remoção de $\mathbf{a}_5$, *mas não de* $\mathbf{a}_1$ *ou de* $\mathbf{a}_2$.

Represente, geometricamente (usando setas), os vetores $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_5$ no plano $\mathbb{R}^2$, e interprete intuitivamente os resultados acima. Ao considerar o item (e), em particular, observe que $\mathbf{a}_1$ e $\mathbf{a}_4$ são colineares.

P4.10. Sejam $\mathbf{u} = \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix}$ e $\mathbf{v} = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$. Verifique que o conjunto $\{\mathbf{u}, \mathbf{v}\}$ é linearmente independente, mas não é uma base de $\mathbb{R}^3$. Encontre um vetor $\mathbf{w} \in \mathbb{R}^3$ de forma que $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ seja uma base de $\mathbb{R}^3$. *Observação*: Há uma infinidade de respostas admissíveis.

P4.11. É possível acrescentar vetores ao conjunto $\left\{ \begin{bmatrix} 6 \\ 0 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix} \right\}$ de forma a obter uma base de $\mathbb{R}^3$? Justifique.

Dica: Veja o exercício P3.27. O teorema 4.19 se aplica nesse caso?

P4.12. Seja $\mathcal{B} = \{\mathbf{a}_1, \ldots, \mathbf{a}_p\}$ uma base do subespaço V de $\mathbb{R}^n$. Suponha que V não seja o subespaço trivial (ou, equivalentemente, que $p \geqslant 1$). Mostre que, para qualquer escalar $\alpha \neq 0$, $\tilde{\mathcal{B}} = \{\alpha\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_p\}$ é também uma base de V. Em particular, V possui uma infinidade de bases distintas.

P4.13. Mostre que $\mathbb{R}^n$ não pode conter um subespaço de dimensão maior do que n, e que o único subespaço de $\mathbb{R}^n$ de dimensão igual a n é o próprio $\mathbb{R}^n$.

Dica: Inspire-se no exemplo 4.28.

P4.14. Verifique que $\mathcal{B} = \left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \end{bmatrix} \right\}$ é uma base de $\mathbb{R}^2$. Em seguida, determine o vetor de coordenadas de cada vetor abaixo com respeito à base $\mathcal{B}$.

(a) $\begin{bmatrix} 5 \\ -1 \end{bmatrix}$ (b) $\begin{bmatrix} 7 \\ 7 \end{bmatrix}$ (c) $\begin{bmatrix} 2 \\ -6 \end{bmatrix}$ (d) $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$

P4.15. Seja $V = \operatorname{Span}\left\{ \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix} \right\}$. Verifique que $\mathcal{B} = \left\{ \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix} \right\}$ é uma base de V. Em seguida, determine se cada vetor abaixo pertence a V. Nos casos afirmativos, calcule o vetor de coordenadas com relação à base $\mathcal{B}$.

(a) $\begin{bmatrix} 4 \\ 12 \\ -7 \end{bmatrix}$ (b) $\begin{bmatrix} 0 \\ -8 \\ 5 \end{bmatrix}$ (c) $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ (d) $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

P4.16. Seja W o plano em $\mathbb{R}^3$ considerado no exemplo 4.37. Já sabemos que $\mathcal{B} = \{\mathbf{a}_1, \mathbf{a}_2\} = \left\{ \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \right\}$ é uma base de W. Seja $\mathbf{u}$ o vetor de W tal que $[\mathbf{u}]_{\mathcal{B}} = \begin{bmatrix} -4 \\ 1 \end{bmatrix}$. Esboce o vetor $\mathbf{u}$ diretamente sobre a figura 4.2. Em seguida, determine as coordenadas de $\mathbf{u}$ com relação à base canônica de $\mathbb{R}^3$.

P4.17. Seja V um subespaço de $\mathbb{R}^n$, e seja $\mathcal{B} = \{\mathbf{a}_1, \ldots, \mathbf{a}_p\}$ uma base de V.

(a) Se $\mathbf{u} \in V$, então o sistema $c_1\mathbf{a}_1 + \cdots + c_p\mathbf{a}_p = \mathbf{u}$ é possível e tem uma única solução. Justifique essa afirmativa.

(b) Se $\mathbf{u} \in \mathbb{R}^n$, mas $\mathbf{u} \notin V$, faz sentido falar das coordenadas de $\mathbf{u}$ com relação a $\mathcal{B}$? Explique. Neste caso, o sistema $c_1\mathbf{a}_1 + \cdots + c_p\mathbf{a}_p = \mathbf{u}$ é possível?

P4.18. Em cada item abaixo, a primeira matriz é linha-equivalente à segunda (neste contexto, o símbolo $\sim$ denota equivalência por operações sobre as linhas). Em cada caso, determine bases para $\operatorname{Col} A$ e $\operatorname{Nuc} A$. Determine também a dimensão desses subespaços.

(a) $A = \begin{bmatrix} 2 & -6 & 8 & 4 \\ -3 & 10 & -9 & -11 \\ 0 & 2 & 6 & -10 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 4 & 2 \\ 0 & 1 & 3 & -5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

(b) $A = \begin{bmatrix} 0 & 0 & 3 & 6 \\ -1 & -2 & 1 & -4 \\ 3 & 6 & 6 & 30 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & -1 & 4 \\ 0 & 0 & 3 & 6 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

(c) $A = \begin{bmatrix} 1 & 4 & 8 & -3 & -7 \\ -1 & 2 & 7 & 3 & 4 \\ -2 & 2 & 9 & 5 & 5 \\ 3 & 6 & 9 & -5 & -2 \end{bmatrix} \sim \begin{bmatrix} 1 & 4 & 8 & 0 & 5 \\ 0 & 2 & 5 & 0 & -1 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

(d) $A = \begin{bmatrix} -1 & 8 & 3 & 3 & -1 \\ 0 & 4 & 8 & 4 & 0 \\ 1 & -2 & 9 & 3 & 1 \\ 2 & -6 & 14 & 7 & 8 \end{bmatrix} \sim \begin{bmatrix} 1 & -6 & 1 & 0 & 3 \\ 0 & 2 & 4 & 2 & 0 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

P4.19. Em cada item a seguir, determine uma base para o subespaço gerado pelos vetores dados.

(a) $\begin{bmatrix} -2 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} -6 \\ -7 \\ 0 \\ -6 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ -2 \\ 1 \end{bmatrix}.$

(b) $\begin{bmatrix} -1 \\ -2 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 1 \\ -3 \end{bmatrix}, \begin{bmatrix} -6 \\ -8 \\ 6 \\ 7 \end{bmatrix}, \begin{bmatrix} 7 \\ -2 \\ -7 \\ 6 \end{bmatrix}.$

P4.20. Seja A a matriz do exercício P4.18(c). Denote as suas colunas por $\mathbf{a}_1$, $\mathbf{a}_2$, ... , $\mathbf{a}_5$.

(a) Verifique que $\mathcal{B} = \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_4\}$ é uma base de $\operatorname{Col} A$. (Essa é, provavelmente, a base que você encontrou no exercício P4.18(c).)

(b) Determine as coordenadas do vetor $\mathbf{a}_5$ com relação à base $\mathcal{B}$.
Dica: Aproveite o trabalho já realizado no exercício P4.18(c), e utilize a proposição 4.43.

P4.21. Reveja o exemplo 4.40 (página 100). Mostre que o vetor $\mathbf{x} = x_2\mathbf{v}_1 + x_4\mathbf{v}_2 + x_5\mathbf{v}_3$ de (4.8) será igual ao vetor zero somente se os pesos x_2, x_4 e x_5 forem todos iguais a zero. (*Dica*: Examine as componentes 2, 4 e 5 dos vetores envolvidos em (4.8).) Conclua que o conjunto $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ é linearmente independente.

P4.22. (a) Convença-se de que cada coluna não-pivô de uma matriz escalonada reduzida é gerada pelas colunas-pivô que estão à esquerda.

Sugestão: Pense em termos de formas escalonadas reduzidas "gerais", como aquelas indicadas em (1.12) (página 12).

(b) Mostre que cada coluna não-pivô de uma matriz *qualquer* é gerada pelas colunas-pivô que estão à esquerda.
Dica: Use o item anterior e a proposição 4.43.

(c) Use o lema 4.17 para concluir que o espaço-coluna de qualquer matriz é gerado por suas colunas-pivô.

P4.23. Observe que os vetores $\mathbf{b}_1$, $\mathbf{b}_2$ e $\mathbf{b}_4$ do exemplo 4.42 são iguais, respectivamente, aos vetores $\mathbf{e}_1$, $\mathbf{e}_2$ e $\mathbf{e}_3$ da base canônica de $\mathbb{R}^4$. Use o exercício P3.27(b) para concluir que esses vetores são linearmente independentes.

P4.24. (a) Seja R uma matriz $n \times m$, na forma escalonada reduzida. Convença-se de que as colunas-pivô de R são iguais a certos elementos da base canônica de $\mathbb{R}^n$. De fato, se R tem p colunas-pivô, então essas colunas são iguais aos vetores $\mathbf{e}_1, \ldots, \mathbf{e}_p$ de $\mathbb{R}^n$.

(b) Conclua, do item anterior, que as colunas-pivô de uma matriz escalonada reduzida são sempre linearmente independentes.

(c) Mostre que as colunas-pivô de uma matriz *qualquer* são linearmente independentes.
Dica: Use novamente a proposição 4.43.

P4.25. Verifique que o vetor $\mathbf{v}_3$ do exemplo 4.50 pode ser escrito como uma combinação linear de $\mathbf{v}_1$ e $\mathbf{v}_2$. Conclua que $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4\}$ é um conjunto gerador para $\operatorname{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_4\}$. (*Dica*: Use o lema 4.17 ou o exercício P4.8.) Verifique também que o conjunto $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4\}$ é linearmente independente. Conclua que esse conjunto é uma base de $\operatorname{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_4\}$, conforme afirmamos no exemplo 4.50.

P4.26. Sejam A e B matrizes linha-equivalentes. Mostre que $\operatorname{Nuc} A = \operatorname{Nuc} B$. Perceba, no entanto, que, em geral, *não vale* $\operatorname{Col} A = \operatorname{Col} B$.

Sugestão: Para a segunda parte, considere os exemplos da subseção 4.5.2.

P4.27. Se o núcleo de uma matriz 6×8 é tridimensional, qual é a dimensão do espaço das colunas?

P4.28. Se o posto de uma matriz 6×8 é igual a dois, qual é a dimensão de seu núcleo?

P4.29. Seja A uma matriz 5×8. A dimensão de $\operatorname{Nuc} A$ pode ser igual a zero? Qual é o menor valor possível para a dimensão de $\operatorname{Nuc} A$? Justifique suas respostas.

P4.30. Determine se cada afirmativa é verdadeira ou falsa. Justifique.

(a) O conjunto vazio é um subespaço.
(b) O subespaço trivial não contém vetor algum.
(c) O subespaço trivial contém exatamente um vetor, portanto não é vazio.

(d) Se $\mathcal{B}$ é uma base de $\mathbb{R}^n$ e V é um subespaço de $\mathbb{R}^n$, então qualquer vetor de V pode ser gerado pelos vetores de $\mathcal{B}$.
(e) Se $\mathcal{B}$ é uma base de $\mathbb{R}^n$, então $\mathcal{B}$ é também uma base de qualquer subespaço de $\mathbb{R}^n$.
(f) Se A e B são matrizes linha-equivalentes, então uma base de $\operatorname{Nuc} A$ é também uma base de $\operatorname{Nuc} B$.
(g) Se A e B são matrizes linha-equivalentes, então uma base de $\operatorname{Col} A$ é também uma base de $\operatorname{Col} B$.
(h) A dimensão do espaço-coluna de uma matriz A é igual ao número de linhas de A.
(i) A dimensão do espaço-coluna de uma matriz A é igual ao número de colunas de A.
(j) Para se obter uma base do espaço-coluna de uma matriz, é recomendável determinar sua forma escalonada *reduzida*.
(k) Para se obter uma base do núcleo de uma matriz, é recomendável determinar sua forma escalonada *reduzida*.

Capítulo 5

Transformações Lineares

5.1 Introdução

Ao considerar a equação matricial $A\mathbf{x} = \mathbf{b}$, podemos mudar o nosso ponto de vista e começar a pensar que o vetor $\mathbf{b}$ foi obtido pela ação da matriz A sobre o vetor $\mathbf{x}$. Sob esse novo ponto de vista, resolver a equação $A\mathbf{x} = \mathbf{b}$ significa encontrar todos os vetores $\mathbf{x} \in \mathbb{R}^n$ que são transformados no vetor $\mathbf{b} \in \mathbb{R}^m$ sob a "ação" da multiplicação por A.

Antes de darmos a definição precisa do que significa uma transformação linear, vamos chamar a atenção para duas propriedades que a multiplicação de uma matriz A por um vetor $\mathbf{u}$ goza. Suponha que:

$$A = \begin{bmatrix} 1 & 3 \\ -3 & 4 \\ 1 & -2 \end{bmatrix}, \mathbf{u} = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \text{ e } \mathbf{v} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

Então,

$$\begin{bmatrix} -6 \\ -21 \\ 9 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ -3 & 4 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} 3 \\ -3 \end{bmatrix} = A(\mathbf{u}+\mathbf{v}) = A(\mathbf{u}) + A(\mathbf{v}) = \begin{bmatrix} -5 \\ -11 \\ 5 \end{bmatrix} + \begin{bmatrix} -1 \\ -10 \\ 4 \end{bmatrix}$$

e

$$\begin{bmatrix} -10 \\ -22 \\ 10 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ -3 & 4 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} 2 \\ -4 \end{bmatrix} = A(2\mathbf{u}) = 2A(\mathbf{u}) = 2 \begin{bmatrix} -5 \\ -11 \\ 10 \end{bmatrix},$$

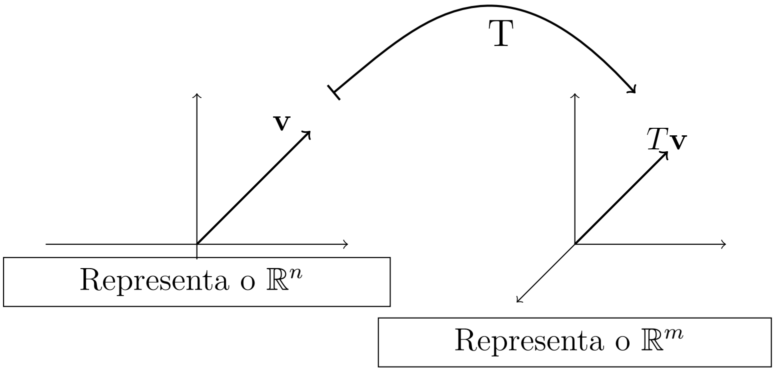
isto é, $A(\mathbf{u}+\mathbf{v}) = A\mathbf{u}+A\mathbf{v}$ e $A(\alpha\mathbf{u}) = \alpha A\mathbf{u}$. E são exatamente essas propriedades que caracterizam a linearidade da função.

Definição 5.1
Uma **Aplicação Linear** (ou **Transformação Linear**) T de $\mathbb{R}^n$ em $\mathbb{R}^m$ é uma função que associa um vetor $\mathbf{v} \in \mathbb{R}^n$ a um vetor $T\mathbf{v} \in \mathbb{R}^m$, tal que para quaisquer $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ e $\alpha \in \mathbb{R}$ tenhamos

(a) $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$;

(b) $T(\alpha\mathbf{u}) = \alpha T(\mathbf{u})$.

Escrevemos $T : \mathbb{R}^n \to \mathbb{R}^m$ e para definir T denotamos $\mathbf{v} \mapsto T\mathbf{v}$.



Assim, $T : \mathbb{R}^n \to \mathbb{R}^m$ é linear se ela "preserva" a soma de vetores e a multiplicação de vetores por escalares, as duas operações básicas do $\mathbb{R}^n$.

Quando tratamos de funções, os termos **domínio**, **contradomínio** e **imagem** aparecem naturalmente. No caso de transformações lineares $T : \mathbb{R}^n \to \mathbb{R}^m$ o domínio é $\mathbb{R}^n$, o contradomínio é $\mathbb{R}^m$ e a imagem é:

$$\mathcal{IM}(T) = \{\mathbf{w} \in R^m : \text{ se existe } \mathbf{v} \in R^n \text{ tal que } T(\mathbf{v}) = \mathbf{w}\} = \{T(\mathbf{v}) : \mathbf{v} \in R^n\}.$$

Exemplo 5.2
Da motivação inicial é claro que se A é uma matriz $m \times n$. Então, A determina uma transformação linear $T : \mathbb{R}^n \to \mathbb{R}^m$, definida por $\mathbf{u} \mapsto T(\mathbf{u}) = A\mathbf{u}$. Sendo mais específicos, digamos que

$$A = \begin{bmatrix} 1 & -4 & 3 \\ 2 & -3 & 0 \end{bmatrix} \text{ e } \mathbf{u} = \begin{bmatrix} x \\ y \\ z \end{bmatrix},$$

então,

$$T(\mathbf{u}) = \begin{bmatrix} 1 & -4 & 3 \\ 2 & -3 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x - 4y + 3z \\ 2x - 3y \end{bmatrix}.$$

Assim, dizemos que $T : \mathbb{R}^3 \to \mathbb{R}^2$, definida por $\begin{bmatrix} x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} x - 4y + 3z \\ 2x - 3y \end{bmatrix}$ provem da matriz A.

Antes de continuarmos veja a seguinte definição.

Definição 5.3
Seja $T : \mathbb{R}^n \to \mathbb{R}^m$ uma transformação linear, no caso em que $m = n$ chamamos a transformação linear (ou aplicação linear) de **operador linear**.

Vamos dar alguns exemplos de aplicações lineares que apresentam os fatos mais marcantes a respeito dessas aplicações. Começando com o seguinte exemplo:

Exemplo 5.4
Seja $T : \mathbb{R}^2 \to \mathbb{R}^2$ dada por

$$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x+3y \\ y \end{bmatrix}$$

Para entendermos o que este operador linear faz, observe que ele toma uma reta e leva em uma reta (aliás, esta propriedade é compartilhada por todas as transformações lineares). Agora considere o quadrado de coordenadas $\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$ e veja que esses pontos são levados em $\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 3 \\ 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 4 \\ 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$, respectivamente. Portanto, este operador linear leva o quadrado em um paralelogramo, esse tipo de transformação linear é conhecido por **cisalhamento**.

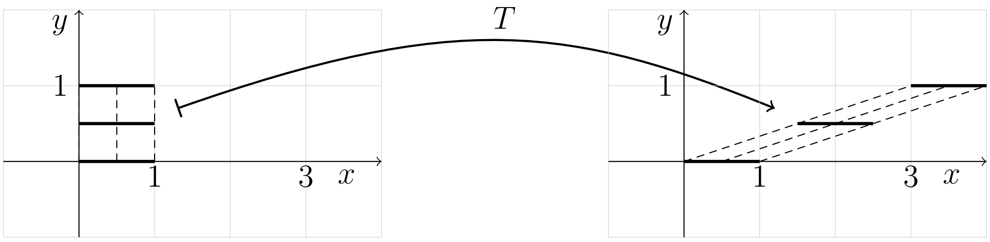


Figura 5.1: Cisalhamento

Exemplo 5.5
Dado um escalar $r > 0$, defina $T : \mathbb{R}^2 \to \mathbb{R}^2$ dada por $\mathbf{v} \mapsto r\mathbf{v}$. Então, chamamos T de **contração** quando $0 < r < 1$ e de **dilatação** quando $r > 1$. Observe que se $\mathbf{u}, \mathbf{v}$ do $\mathbb{R}^2$ e escalares α, β, então

$$\begin{aligned} T(\alpha\mathbf{u} + \beta\mathbf{v}) &= r(\alpha\mathbf{u} + \beta\mathbf{v}) \\ &= r\alpha\mathbf{u} + r\beta\mathbf{v} \\ &= \alpha(r\mathbf{u}) + \beta(r\mathbf{v}) \\ &= \alpha T(\mathbf{u}) + \beta T(\mathbf{v}). \end{aligned} \tag{5.1}$$

Portanto, T é um operador linear, pois se tomarmos $\alpha = \beta = 1$ na equação anterior temos a primeira condição para T ser uma aplicação linear e se fizermos $\beta = 0$ temos a segunda condição.

Exemplo 5.6
Considere a aplicação $T : \mathbb{R}^2 \to \mathbb{R}^2$ dada por $\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} x \\ y^2 \end{bmatrix}$. Vamos mostrar que T não é um operador linear. Para isso considere os vetores: $\left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$. Portanto:

$$T\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = T\left(\begin{bmatrix} 2 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$$

e

$$T\left(\begin{bmatrix}1\\2\end{bmatrix}\right)+T\left(\begin{bmatrix}1\\0\end{bmatrix}\right)=\begin{bmatrix}1\\4\end{bmatrix}+\begin{bmatrix}1\\0\end{bmatrix}=\begin{bmatrix}3\\4\end{bmatrix}.$$

Se T fosse um operador linear deveria ocorrer a igualdade, isto é,

$$T\left(\begin{bmatrix}1\\2\end{bmatrix}+\begin{bmatrix}1\\0\end{bmatrix}\right)=T\left(\begin{bmatrix}1\\2\end{bmatrix}\right)+T\left(\begin{bmatrix}1\\0\end{bmatrix}\right).$$

5.2 Propriedades de uma Transformação Linear e sua Matriz

Se $T:\mathbb{R}^n\to\mathbb{R}^m$ é uma transformação linear, então $T(\mathbf{0})=\mathbf{0}$. De fato, seja $\mathbf{u}\in\mathbb{R}^n$ então $T(\mathbf{0})=T(0\mathbf{u})=0T(\mathbf{u})=\mathbf{0}$.

Considere novamente a transformação linear $T:\mathbb{R}^n\to\mathbb{R}^m$ e suponha que $\mathbf{u}_1,\mathbf{u}_2,\dots,\mathbf{u}_p$ são vetores de $\mathbb{R}^n$ e seja $\mathbf{u}=\alpha_1\mathbf{u}_1+\alpha_2\mathbf{u}_2+\cdots+\alpha_p\mathbf{u}_p$ uma combinação linear destes vetores logo:

$$\begin{aligned}T(\mathbf{u})&=T(\alpha_1\mathbf{u}_1+\alpha_2\mathbf{u}_2+\cdots+\alpha_p\mathbf{u}_p)\\&=T(\alpha_1\mathbf{u}_1)+T(\alpha_2\mathbf{u}_2)+\cdots+T(\alpha_p\mathbf{u}_p)\\&=\alpha_1T(\mathbf{u}_1)+\alpha_2T(\mathbf{u}_2)+\cdots+\alpha_pT(\mathbf{u}_p).\end{aligned}\tag{5.2}$$

Isso nos sugere duas consequências importantes: A primeira consequência é a de que se soubermos como uma aplicação linear age sobre uma base de $\mathbb{R}^n$, podemos determinar a ação da transformação linear no restante dos vetores por aplicar a fórmula anterior; a segunda consequência é a de que podemos associar uma matriz a uma transformação linear, da seguinte forma: suponha que $\mathbf{e}_1,\mathbf{e}_2,\dots,\mathbf{e}_n$ é a base canônica do $\mathbb{R}^n$. Então dado qualquer vetor $\mathbf{u}=\begin{bmatrix}x_1\\\vdots\\x_n\end{bmatrix}=\sum_{i=1}^n x_i\mathbf{e}_i$ e daí,

$$T\left(\begin{bmatrix}x_1\\\vdots\\x_n\end{bmatrix}\right)=\sum_{i=1}^n x_iT(\mathbf{e}_i)$$

e $T(\mathbf{e}_i)$ são vetores de $\mathbb{R}^m$, isto é, $T(\mathbf{u})$ é uma combinação linear $T(\mathbf{e}_i)\in\mathbb{R}^m$. Escrevendo isto em termos da multiplicação de matrizes temos:

$$\begin{bmatrix}T(\mathbf{e}_1) & T(\mathbf{e}_2) & \cdots & T(\mathbf{e}_n)\end{bmatrix}\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}.$$

Portanto, a transformação linear T fica completamente determinada pela matriz $\begin{bmatrix}T(\mathbf{e}_1) & T(\mathbf{e}_2) & \cdots & T(\mathbf{e}_n)\end{bmatrix}_{m\times n}$ e chamamos esta matriz de **matriz da aplicação linear na base canônica**. Mais para frente esse nome fará mais sentido. No momento vamos dar alguns exemplos.

Exemplo 5.7
Vamos retornar ao exemplo 5.5, em que $T : \mathbb{R}^2 \to \mathbb{R}^2$, que toma $\mathbf{u} \in \mathbb{R}^2 \mapsto 3\mathbf{u} \in \mathbb{R}^2$, que tal como havíamos comentado é uma dilatação e é um operador linear. Vamos determinar qual seria a matriz desse operador linear com respeito à base canônica. Seja $\mathbf{e}_1 = \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right], \mathbf{e}_2 = \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ a base canônica de $\mathbb{R}^2$, logo, $T(\mathbf{e}_1) = \left[\begin{smallmatrix}3\\0\end{smallmatrix}\right]$ e $T(\mathbf{e}_2) = \left[\begin{smallmatrix}0\\3\end{smallmatrix}\right]$ e, portanto, a matriz fica:

$$A = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}.$$

Exemplo 5.8
Considere a transformação linear $R : \mathbb{R}^2 \mapsto \mathbb{R}^2$, que para cada vetor $\mathbf{v}$ retorna um vetor $R(\mathbf{v})$, obtido por fazer uma rotação de θ radianos em torno da origem no sentido anti-horário. A próxima figura nos fornece uma demonstração geométrica de que $R(\mathbf{u} + \mathbf{v}) = R(\mathbf{u}) + R(\mathbf{v})$.

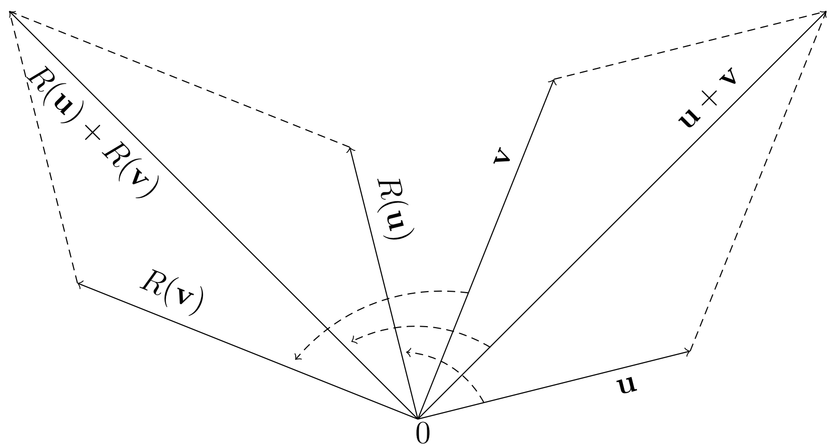


Figura 5.2: Rotação de vetores

Com um desenho similar se prova que $R(\alpha\mathbf{v}) = \alpha R(\mathbf{v})$. Portanto, esta aplicação é linear. Vamos determinar a sua matriz na base canônica. Para isto, vamos investigar onde os vetores $\mathbf{e}_1 = \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right], \mathbf{e}_2 = \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ são levados. Na figura abaixo vemos que: $R(\mathbf{e}_1) = \begin{bmatrix} \cos\theta \\ \operatorname{sen}\theta \end{bmatrix}$, $R(\mathbf{e}_2) = \begin{bmatrix} -\operatorname{sen}\theta \\ \cos\theta \end{bmatrix}$ e a matriz na base canônica desta transformação é:

$$A = \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix}.$$

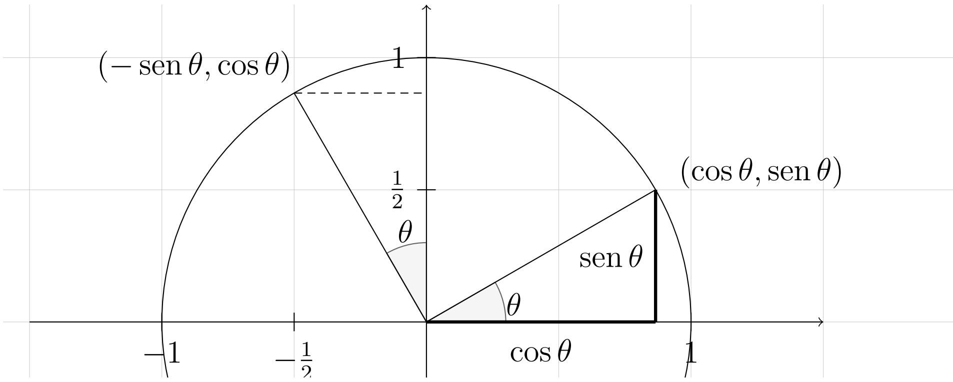


Figura 5.3: Rotação de um ângulo θ

O próximo exemplo trata da projeção ortogonal sobre o eixo x e o outro da reflexão em torno desse eixo. Estes são exemplos de duas importantes famílias de aplicações lineares as projeções e as reflexões.

Exemplo 5.9
Considere a aplicação $P : \mathbb{R}^2 \to \mathbb{R}^2$ definida por $\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] \mapsto \left[\begin{smallmatrix} x \\ 0 \end{smallmatrix}\right]$. É possível essa verificação porque a mesma faz um gráfico em que essa aplicação é linear, uma vez que ela pega um vetor $\mathbf{u}$ qualquer e projeta perpendicularmente sobre o eixo x. Essa aplicação é conhecida como projeção ortogonal sobre o eixo do x.

Agora a reflexão em torno do eixo x.

Exemplo 5.10
A reflexão $S : \mathbb{R}^2 \to \mathbb{R}^2$ do vetor $\mathbf{u}$ sobre o eixo x toma o vetor $\mathbf{u}$ e retorna o vetor $S\mathbf{u}$ do lado oposto desse eixo, de tal maneira que o segmento de reta determinado por $\mathbf{u}$ e $S\mathbf{u}$ é perpendicular ao eixo x e o eixo corta este segmento no ponto médio.

Observe que o vetor $\mathbf{u} + S\mathbf{u}$ está sobre o eixo x, uma vez que o eixo é exatamente a diagonal do paralelogramo determinado pelos vetores $\mathbf{u}$ e $S\mathbf{u}$. Além disso, o vetor $\mathbf{u}+S\mathbf{u}$ é igual a duas vezes a projeção de $\mathbf{u}$ sobre o eixo x. Portanto:

$$2P\mathbf{u} = \mathbf{u} + S\mathbf{u} \Leftrightarrow S\mathbf{u} = 2P\mathbf{u} - \mathbf{u}.$$

Logo,

$$S\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = 2P\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) - \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2x \\ 0 \end{bmatrix} - \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ -y \end{bmatrix}.$$

5.2.1 Composição e Aplicação Invertível

Considere duas funções (ou aplicações) $f : A \to B$ e $g : B \to C$. A **composição**, denotada por $f \circ g$, é a aplicação $f \circ g : A \to C$, definida por:

$$(f \circ g)(a) = f(g(a)),$$

isto é, primeiro aplicamos f em $a \in A$ e depois aplicamos g em $f(a)$ para obter $g(f(a)) \in C$.

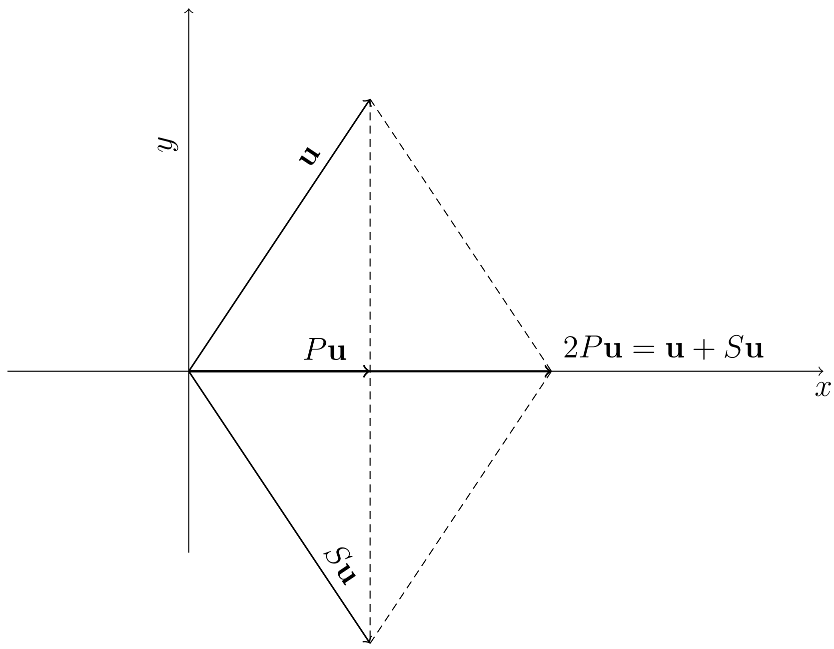


Figura 5.4: Reflexão em torno do eixo x

Exemplo 5.11

Considere $f : \mathbb{R} \to \mathbb{R}$ definida por $f(x) = 2x^2 - 3x$ e $g : \mathbb{R} \to \mathbb{R}$ definida por $g(x) = x^2$ então

$$(f \circ g)(x) = f(g(x)) = f(x^2) = 2(x^2)^2 - 3(x^2) = 2x^4 - 3x^2.$$

Se A é um conjunto não-vazio, a aplicação $f : A \to A$ definida por $f(a) = a$ para todo $a \in A$ é chamada de **aplicação identidade** sendo, geralmente, denotada por I_A ou simplesmente por I. Assim, $I(a) = a$ para todo $a \in A$.

Seja $f : A \to B$ uma aplicação. Dizemos que $g : B \to A$ é a **aplicação inversa** de f, denotada por $g = f^{-1}$, se

$$f \circ g = I_B \quad \text{e} \quad g \circ f = I_A.$$

Exemplo 5.12

Considere o operador linear $T : \mathbb{R}^2 \to \mathbb{R}^2$ definida por $\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} 2x + 5y \\ 3x + 7y \end{bmatrix}$. Verifique se S definida por $\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} -7x + 5y \\ 3x - 2y \end{bmatrix}$ é a inversa de T.

Observe que

$$S\left(T\begin{bmatrix} x \\ y \end{bmatrix}\right) = S\begin{bmatrix} 2x + 5y \\ 3x + 7y \end{bmatrix} = \begin{bmatrix} -7(2x + 5y) + 5(3x + 7y) \\ 3(2x + 5y) - 2(3x + 7y) \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}.$$

e

$$T\left(S\begin{bmatrix} x \\ y \end{bmatrix}\right) = T\begin{bmatrix} -7x + 5y \\ 3x - 2y \end{bmatrix} = \begin{bmatrix} 2(-7x + 5y) + 5(3x - 2y) \\ 3(-7x + 5y) + 7(3x - 2y) \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}.$$

Portanto, $S = T^{-1}$.

5.2.2 Aplicação Linear Injetora e Sobrejetora

As aplicações lineares formam uma classe bastante especial de funções na qual é fácil decidir se uma certa aplicação é invertível ou não. Inicialmente, vamos lembrar que uma aplicação $f : A \to B$ tem inversa se, e somente se, f for uma aplicação injetora e sobrejetora. Por isso, para saber se uma certa aplicação é invertível, precisamos estudar a fim de constatar se a aplicação é injetora e sobrejetora.

Dizemos que a transformação linear $f : A \to B$ é injetora se para cada $a \neq b$ em A tivermos que $f(a) \neq f(b)$ em B. Apesar dessa definição ser mais natural, é comum utilizarmos a seguinte forma equivalente: $f : A \to B$ é injetora se sempre que $f(a) = f(b)$, quando pudermos mostrar que $a = b$.

No caso da aplicação ser uma aplicação linear, digamos $T : \mathbb{R}^n \to \mathbb{R}^m$, podemos substituir este critério pelo seguinte: se $T(\mathbf{u}) = \mathbf{0}$ em $\mathbb{R}^m$, então $\mathbf{u} = \mathbf{0}$.

De fato, vamos admitir que $T : \mathbb{R}^n \to \mathbb{R}^m$ é uma aplicação linear e injetora e como sabemos que o vetor nulo de $\mathbb{R}^n$ é sempre levado no vetor nulo de $\mathbb{R}^m$, então se

$$T(\mathbf{u}) = \mathbf{0} = T(\mathbf{0}) \Rightarrow \mathbf{u} = \mathbf{0} \in \mathbb{R}^n$$

e, reciprocamente se admitir que sempre ocorre $T(\mathbf{u}) = \mathbf{0} \Rightarrow \mathbf{u} = \mathbf{0}$, então se

$$T(\mathbf{u}) = T(\mathbf{v}) \Leftrightarrow T(\mathbf{u}) - T(\mathbf{v}) = \mathbf{0} \Leftrightarrow T(\mathbf{u} - \mathbf{v}) = \mathbf{0},$$

logo $\mathbf{u} - \mathbf{v} = \mathbf{0}$. Portanto, $\mathbf{u} = \mathbf{v}$. Daí que T é injetora.

Por isso, no estudo da injetividade de uma aplicação linear é muito útil a introdução do seguinte conceito:

Seja $T : \mathbb{R}^n \to \mathbb{R}^m$ uma transformação linear o subconjunto dos vetores de $\mathbb{R}^n$ que são levados por T no vetor nulo de $\mathbb{R}^m$ é chamado de **núcleo** de T e é denotado por:

$$\mathcal{N}(T) = \{\mathbf{u} \in \mathbb{R}^n : T(\mathbf{u}) = \mathbf{0}\}.$$

Segue do que foi discutido anteriormente que $\mathcal{N}(T) = \{\mathbf{0}\}$ se, e somente se, T é injetora.

Exemplo 5.13
Seja $T : \mathbb{R}^2 \to \mathbb{R}$ definida por $\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] \mapsto x - y$. Ao calcularmos $\mathcal{N}(T)$ obtemos $\mathcal{N}(T) = \left\{\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] \in \mathbb{R}^2 : x - y = 0\right\}$, que coincide com a reta $y = x$, ou ainda, $\mathcal{N}(T) = \operatorname{Span}\left\{\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]\right\}$.

Dizemos que $f : A \to B$ é **sobrejetora** se, para todo $b \in B$, existe $a \in A$ tal que $f(a) = b$.

Reescrevendo esse critério, para aplicações lineares, obtemos: uma aplicação linear $T : \mathbb{R}^n \to \mathbb{R}^m$, é **sobrejetora** se para cada vetor $\mathbf{v} \in \mathbb{R}^m$ é imagem de algum vetor $\mathbf{u} \in \mathbb{R}^n$. Isso é equivalente a pedir que $\mathcal{IM}(T) = \mathbb{R}^m$.

Exemplo 5.14

Seja $T : \mathbb{R}^3 \to \mathbb{R}^3$ definida por $\left[\begin{smallmatrix} -y \\ 3x \\ 0 \end{smallmatrix}\right]$. Então, a imagem de T

$$\begin{aligned}
\mathcal{IM}(T) &= \left\{ \begin{bmatrix} -y \\ 3x \\ 0 \end{bmatrix} : x,\, y \in \mathbb{R} \right\} = \left\{ x \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix} + y \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} : x,\, y \in \mathbb{R} \right\} \\
&= \operatorname{Span}\left\{ \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} \right\}.
\end{aligned}$$

O núcleo de T é dado pelos vetores $\mathbf{u} = \left[\begin{smallmatrix} x \\ y \\ z \end{smallmatrix}\right]$, tais que:

$$T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} \right) = \begin{bmatrix} -y \\ 3x \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Portanto, $x = y = 0$ e daí

$$\mathcal{N}(T) = \operatorname{Span}\left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

5.3 Álgebra Matricial

Vamos introduzir operações sobre as matrizes e depois mostrar que essas operações provêm das operações sobre as transformações lineares. Para começar, vamos definir o que significa dizer que a matriz $A = [a_{ij}]$ é igual à matriz $B = [b_{ij}]$, isso significa que ambas as matrizes são $m \times n$ e que as entradas correspondentes são iguais $a_{ij} = b_{ij}$. Quando isso acontece denotamos por $A = B$.

Dado um escalar $r \in \mathbb{R}$ e a matriz $A = [a_{ij}]$, definimos o produto de r por A, por ser a matriz $r \cdot A = [ra_{ij}]$, isto é, cada uma das entradas da matriz é multiplicada pelo escalar r. E dado as matrizes $A = [a_{ij}]$ e $B = [b_{ij}]$ de ordem $m \times n$, definimos a soma das matrizes por ser uma outra matriz, também de ordem $m \times n$, dada por

$$A + B = [a_{ij} + b_{ij}],$$

isto é, entrada da matriz $A+B$ é a soma das entradas correspondentes de A e B.

Essas operações têm as seguintes propriedades:

a. $A + B = B + A$	d. $r(A + B) = rA + rB$
b. $(A + B) + C = A + (B + C)$	e. $(r + s)A = rA + sB$
c. $(A + 0) = A$	f. $r(sA) = (rs)A$

5.3.1 A Transposta de uma Matriz

Dada a matriz A de ordem $m \times n$, a **transposta** de A é a matriz $n \times m$, denotada por A^t, cujas colunas são formadas pelas linhas correspondentes de A.

Exemplo 5.15

$$A = \begin{bmatrix} -5 & 4 & 2 \\ 0 & 2 & 1 \end{bmatrix} \quad \text{e} \quad B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Então,

$$A^t = \begin{bmatrix} -5 & 0 \\ 4 & 2 \\ 2 & 1 \end{bmatrix} \quad \text{e} \quad B^t = \begin{bmatrix} a & c \\ b & d \end{bmatrix}.$$

Se A e B são matrizes do mesmo tipo e $r \in \mathbb{R}$, logo valem as seguintes propriedades:

a. $(A^t)^t = A$;
b. $(A+B)^t = A^t + B^t$;
c. $(rA)^t = rA^t$;

Denotaremos por $\mathbf{M}(m \times n)$ ao conjunto de todas as matrizes de ordem $m \times n$.

5.3.2 Multiplicação de Matrizes

Vamos introduzir a multiplicação de matrizes que tem uma definição um tanto complicada. Mais para frente veremos que definimos a multiplicação para que a mesma possa ser utilizada na composição de transformações lineares.

Considere as matrizes $A = [a_{ij}] \in \mathbf{M}(m \times n)$ e $B = [b_{ij}] \in M(n \times k)$ definimos a matriz $AB = [d_{ij}] \in M(m \times k)$ por

$$d_{ij} = \sum_{l=1}^{n} a_{il} b_{lj}.$$

Parafraseando, o elemento da i-ésima linha e j-ésima coluna da matriz AB é a soma dos produtos dos elementos correspondentes da i-ésima linha de A e da j-ésima coluna de B. Observe, em primeiro lugar que o tamanho das linhas de A é igual ao tamanho das colunas de B e que a quantidade de linhas (colunas) da matriz AB é igual a quantidade de linhas de A (colunas de B).

Exemplo 5.16
Considere A uma matriz 2×2 e B uma matriz 2×3 então AB é uma matriz 2×3 e a entrada 12 é

$$AB = \begin{bmatrix} 3 & 2 \\ -4 & 2 \end{bmatrix} \begin{bmatrix} 4 & 1 & 6 \\ -1 & -5 & 3 \end{bmatrix} = \begin{bmatrix} \square & 3(1) + 2(-5) & \square \\ \square & \square & \square \end{bmatrix} = \begin{bmatrix} \square & -7 & \square \\ \square & \square & \square \end{bmatrix}.$$

Também podemos encarar a multiplicação da matriz como multiplicação de vetores, veja exemplo abaixo

Exemplo 5.17
Considere as matrizes

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} \text{ e } B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 \end{bmatrix},$$

onde $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ e $\mathbf{w}_1, \mathbf{w}_2$ são os vetores colunas de A e B, respectivamente. Então,

$$\begin{aligned} AB &= \begin{bmatrix} b_{11}a_{11} + b_{21}a_{12} + b_{31}a_{13} & b_{12}a_{11} + b_{22}a_{12} + b_{32}a_{13} \\ b_{11}a_{21} + b_{21}a_{22} + b_{31}a_{23} & b_{12}a_{21} + b_{22}a_{22} + b_{32}a_{23} \\ b_{11}a_{31} + b_{21}a_{32} + b_{31}a_{33} & b_{12}a_{31} + b_{22}a_{32} + b_{32}a_{33} \end{bmatrix} \\ &= \begin{bmatrix} b_{11}\mathbf{v}_1 + b_{21}\mathbf{v}_2 + b_{31}\mathbf{v}_3 & b_{12}\mathbf{v}_1 + b_{22}\mathbf{v}_2 + b_{32}\mathbf{v}_3 \end{bmatrix} \\ &= \begin{bmatrix} A\mathbf{w}_1 & A\mathbf{w}_2 \end{bmatrix}. \end{aligned}$$

Como podemos ver o resultado do produto AB pode ser visto como vetores colunas, onde cada coluna é uma combinação linear dos vetores colunas da matriz A, com os coeficientes da coluna correspondente da matriz B, e também, os vetores colunas de AB podem ser visto como o produto de A pelos vetores colunas correspondentes de B. Claramente isso se generaliza para matrizes de qualquer ordem em que o produto faça sentido.

A matriz $0 = [0] \in M(m \times n)$ é chamada de **matriz nula**. A matriz nula possui a seguinte propriedade: para quaisquer outras matrizes $A \in M(k \times m)$ e $B \in M(n \times k)$ temos

$$A0 = 0 \in M(k \times n) \text{ e } 0B = 0 \in M(m \times k).$$

Outra matriz que desempenha um papel importante é a matriz quadrada $I = [\delta_{ij}]$ onde

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

Esta matriz tem a seguinte propriedade: qualquer que seja a matriz $A = [a_{ij}] \in M_n$ temos

$$A \cdot I = A \text{ e } I \cdot A = A,$$

chamaremos tal matriz de **matriz identidade**.

Com respeito a operação de multiplicação de matrizes ela possui as seguintes propriedades: suponha que $A \in \mathbf{M}(m \times n)$, e sejam B e C matrizes do tipo adequado para que as operações de soma e produto sejam possíveis

a. $A(BC) = (AB)C$	Associatividade da multiplicação;
b. $A(B + C) = AB + AC$	propriedade distributiva à esquerda;
b'. $(B + C)A = BA + CA$	propriedade distributiva à direita;
c. $r(AB) = (rA)B = A(rB)$	r é um escalar;
d. $I_mA = A = AI_m$	onde I_m é a matriz identidade;
e. $(AB)^t = B^tA^t$.	

Exemplo 5.18
Em geral a propriedade comutativa não é verdadeira para as matrizes. Para perceber isso considere:

$$A = \begin{bmatrix} 0 & 2 \\ -3 & 1 \end{bmatrix} \text{ e } B = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix},$$

logo,

$$AB = \begin{bmatrix} -4 & 2 \\ -5 & 1 \end{bmatrix} \text{ e } BA = \begin{bmatrix} 0 & 2 \\ -3 & -3 \end{bmatrix} \text{ e, claramente, } AB \neq BA.$$

Diremos que uma matriz $B \in \mathbf{M}(n \times m)$ é o **inverso à direita** de $A \in \mathbf{M}(m \times n)$ se $AB = I_m$ e diremos que a matriz $C \in \mathbf{M}(n \times m)$ é o **inverso à esquerda** de A se $CA = I_n$.

No caso em que a matriz $A \in \mathbf{M}(m \times n)$ possua inverso à direita B e à esquerda C então:

$$B = I_n B = (CA)B = C(AB) = CI_m = C,$$

e isso significa que $B = C$, em particular $m = n$ e a matriz A é quadrada. Logo, se A possui inverso à esquerda e à direita ao mesmo tempo, eles são os mesmos, além disso, pelo mesmo argumento usado anteriormente, se A possui inversa à direita e à esquerda ao mesmo tempo, então esta inversa é única e chamamos esta matriz de **matriz inversa** de A e a denotamos por A^{-1} e, nesse caso, diremos simplesmente que A é invertível.

Exemplo 5.19
Vamos começar por considerar a matriz

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

Queremos determinar uma matriz $B = \begin{bmatrix} x & y \\ z & w \end{bmatrix}$, tal que $AB = I_2$, mas isso implica no seguinte:

$$AB = \begin{bmatrix} 2 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} x & y \\ z & w \end{bmatrix} = \begin{bmatrix} 2x+z & 2y+w \\ 3x+z & 3y+w \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

e daí, precisamos resolver o seguinte sistema:

$$\begin{cases} 2x & & +z & & = & 1 \\ & 2y & & +w & = & 0 \\ 3x & & +z & & = & 0 \\ & 3y & & +w & = & 1 \end{cases}.$$

Nesse caso, resolvendo por escalonamento, obtemos $x = -1, y = 1, z = 3$ e $w = -2$ e portanto, $B = \begin{bmatrix} -1 & 1 \\ 3 & -2 \end{bmatrix}$, além disso, se fizermos BA também obtemos a matriz identidade e, portanto, $B = A^{-1}$.

Uma matriz $A \in \mathbf{M}(m \times n)$ pode admitir mais de uma inversa à direita.

Exemplo 5.20
Se

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

considere para quaisquer que sejam $a, b \in \mathbb{R}$ a seguinte matriz:

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

daí,

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

Por outro lado, uma matriz $A \in \mathbf{M}(m \times n)$ pode admitir mais de uma inversa à esquerda.

Exemplo 5.21
Considere a matriz:

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

e considere quaisquer $a, b \in \mathbb{R}$ e $C = \begin{bmatrix} 1 & 0 & a \\ 0 & 1 & b \end{bmatrix}$, logo,

$$CA = \begin{bmatrix} 1 & 0 & a \\ 0 & 1 & b \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 3 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Vamos demonstrar o seguinte teorema que nos fornece algumas propriedades úteis a respeito da inversa de uma matriz A.

Teorema 5.22
a) Se $A \in \mathbf{M}_n$ é invertível, então a sua inversa A^{-1} é única;

b) Se A é uma matriz invertível, então A^{-1} também o é e $(A^{-1})^{-1} = A$;

c) Se $A, B \in \mathbf{M}_n$ são matrizes invertíveis, então AB também é invertível e

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

Demonstração: Como a) já foi estabelecida, vamos demonstrar b). Dizer que A^{-1} é invertível significa dizer que existe uma matriz C, tal que

$$CA^{-1} = I \text{ e } A^{-1}C = I.$$

Mas, A satisfaz estas equações e como a inversa é única pela letra a), segue que $(A^{-1})^{-1} = A$. Já na letra c) observe que como A e B são invertíveis existem únicas A^{-1} e B^{-1} e veja que:

$$\begin{aligned} (B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I \text{ e} \\ (AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I, \end{aligned} \tag{5.3}$$

portanto, AB é invertível e a sua inversa é $B^{-1}A^{-1}$. □

5.4 Matrizes Elementares

O objetivo dessa seção é introduzir as matrizes elementares e com elas descrever um processo que permite obter a matriz inversa, isso é claro, nos casos em que a matriz admite uma inversa. Gostaria de chamar a atenção que as matrizes elementares também podem ser usadas na criação de algoritmos que permitem manipular matrizes, dito isso, vamos ao que nos interessa. Iniciamos lembrando que existem apenas três operações elementares que podemos usar no processo de escalonamento, que são:

(a) $\ell_i \to \ell_i + \alpha\ell_j$;

(b) $\ell_i \to \alpha\ell_i$;

(c) $\ell_i \leftrightarrow \ell_j$.

Diremos que E é uma **matriz elementar** se E pode ser obtido da matriz identidade I por aplicar apenas umas das três operações elementares.

Exemplo 5.23
Veja que a matriz $E = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}$ é uma matriz elementar pois pode ser obtida por aplicar $\ell_2 \to \ell_2 - 2\ell_1$ na matriz identidade.

As matrizes elementares têm inúmeras aplicações, em particular, nós podemos substituir as operações elementares sobre as linhas (as colunas), por multiplicar a matrizes elementares à esquerda (à direita) da matriz que queremos que sofra a operação elementar. Veja o próximo exemplo.

Exemplo 5.24
Digamos que gostaríamos de aplicar a operação elementar $\ell_2 \to \ell_2 - 2\ell_1$ na matriz $A = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 4 & 1 \end{bmatrix}$, mas ao invés de fazer isso faça:

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

É fácil de ver que o resultado é o mesmo. Isso nos diz que podemos substituir o procedimento de aplicar uma operação elementar em uma matriz por multiplicar a mesma, à esquerda, pela matriz elementar obtida da matriz identidade por aplicar mesma operação elementar.

Olhando o resultado, vemos que o próximo passo, para escalonar a matriz seria aplicar $\ell_2 \to -\frac{1}{2}\ell_2$ na matriz obtida acima. Então vamos multiplicar por $E' = \begin{bmatrix} 1 & 0 \\ 0 & -\frac{1}{2} \end{bmatrix}$ e teremos

$$E'EA = E'(EA) = \begin{bmatrix} 1 & 0 \\ 0 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 3 & 2 \\ 0 & -2 & -3 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 2 \\ 0 & 1 & \frac{3}{2} \end{bmatrix}.$$

Para demonstrar o resultado principal enunciado no teorema a seguir vamos necessitar do seguinte lema:

Lema 5.25
Se E é uma matriz elementar, então E é invertível.

Demonstração: De fato se E é uma matriz elementar, então podemos determinar uma matriz elementar E' obtida por fazer a operação inversa que a efetuada para obter E. Logo, $E'E = I$. $\square$

O resultado principal fica.

Teorema 5.26
Seja A uma matriz quadrada de ordem n. Então, A é invertível se, e somente se, a sua forma escalonada reduzida for a matriz identidade.

Demonstração. Suponha que A é invertível, isto é, existe uma matriz B $n \times n$, tal que $AB = I = BA$. Vamos começar analisando o sistema $A\mathbf{x} = \mathbf{0}$ para mostrar que a única solução possível é $\mathbf{x} = \mathbf{0}$. De fato, multiplicando a equação $A\mathbf{x} = \mathbf{0}$ por B temos que $0 = B \cdot 0 = B(A\mathbf{x}) = (BA)\mathbf{x} = I\mathbf{x} = \mathbf{x}$. Por outro lado, se resolvermos o sistema $A\mathbf{x} = \mathbf{0}$ através de um escalonamento de linhas, obtemos a matriz R, que está escalonada de forma reduzida e é linha equivalente A, ou seja, $R = E_k \cdots E_1 A$, onde $E_1, \ldots, E_k$ são matrizes elementares. Mas o sistema $R\mathbf{x} = \mathbf{0}$ é claramente equivalente ao sistema, $A\mathbf{x} = \mathbf{0}$, portanto tem a mesma e única solução $\mathbf{x} = \mathbf{0}$, e como R está na forma escalonada reduzida, a única possibilidade é de $R = I$. Logo, se A é invertível então ela é linha equivalente a matriz identidade. Reciprocamente, se A é linha equivalente a matriz identidade, isto é, $I = E_k \cdots E_1 A$, e multiplicando por $E_k^{-1}, E_{k-1}^{-1}, \ldots, E_1^{-1} I$, obtemos que $A = E_1^{-1} E_2^{-1} \cdots E_k^{-1}$, ou seja, A é um produto de matrizes elementares e portanto invertível. $\square$

Analisando a demonstração do teorema 5.26 podemos vislumbrar um método para obtermos a inversa de uma matriz. Suponha que A é uma matriz invertível, isto é, existe um número finito de matrizes elementares $E_1, \ldots, E_k$, tais que:

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

Isso mostra que $E_k E_{k-1} \cdots E_2 E_1 I = A^{-1}$. Isto é, se aplicarmos as mesmas operações, e na mesma ordem, necessárias para levar a matriz A a sua forma escalonada reduzida, na matriz identidade, obtemos a matriz inversa de A. Isso nos motiva a definir um algoritmo para obtermos a inversa de uma matriz. Para isso basta considerar uma nova matriz $B = [A : I]$, por acrescentar a matriz identidade a direita de A então, se escalonarmos a matriz A, para que se torne a matriz identidade, segue que B se torna $[I : C]$ e então, C será a matriz inversa de A.

Exemplo 5.27
Vamos usar esse processo para obter a inversa da matriz

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

Para isso considere a matriz:

$$[A|I] = \left[\begin{array}{rrr|rrr} 1 & -1 & 2 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 1 & 0 \\ 2 & -3 & 1 & 0 & 0 & 1 \end{array}\right].$$

Vamos escaloná-la. Comece por fazer $\ell_2 \to \ell_2 + \ell_1$ e $\ell_3 \to \ell_3 - 2\ell_1$ e obtemos

$$\left[\begin{array}{rrr|rrr} 1 & -1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 & 1 & 0 \\ 0 & -1 & -3 & -2 & 0 & 1 \end{array}\right] \xrightarrow{\ell_2 \leftrightarrow \ell_3} \left[\begin{array}{rrr|rrr} 1 & -1 & 2 & 1 & 0 & 0 \\ 0 & -1 & -3 & -2 & 0 & 1 \\ 0 & 0 & 2 & 1 & 1 & 0 \end{array}\right] \xrightarrow{\ell_1 \to \ell_1 - \ell_2}$$

$$\left[\begin{array}{rrr|rrr} 1 & 0 & 5 & 3 & 0 & -1 \\ 0 & -1 & -3 & -2 & 0 & 1 \\ 0 & 0 & 2 & 1 & 1 & 0 \end{array}\right] \xrightarrow[\ell_2 \to \ell_2 + \frac{3}{2}\ell_3]{\ell_1 \to \ell_1 - \frac{5}{2}\ell_3} \left[\begin{array}{rrr|rrr} 1 & 0 & 0 & 1/2 & -5/2 & -1 \\ 0 & -1 & 0 & -1/2 & 3/2 & 1 \\ 0 & 0 & 2 & 1 & 1 & 0 \end{array}\right]$$

$$\xrightarrow[\ell_3 \to \frac{1}{2}\ell_3]{\ell_2 \to -\ell_2} \left[\begin{array}{rrr|rrr} 1 & 0 & 0 & 1/2 & -5/2 & -1 \\ 0 & 1 & 0 & 1/2 & -3/2 & -1 \\ 0 & 0 & 1 & 1/2 & 1/2 & 0 \end{array}\right]$$

$$\text{e, portanto, } A^{-1} = \begin{bmatrix} 1/2 & -5/2 & -1 \\ 1/2 & -3/2 & -1 \\ 1/2 & 1/2 & 0 \end{bmatrix}.$$

5.5 Núcleo e Imagem

Nesta seção vamos conectar os conceitos da dimensão da imagem de uma transformação linear com a dimensão do seu núcleo, os quais nos darão um invariante importante para as aplicações lineares. Vamos começar com o seguinte teorema:

Teorema 5.28
Seja $T : \mathbb{R}^n \to \mathbb{R}^m$ uma transformação linear. Então o núcleo de T é um subespaço vetorial de $\mathbb{R}^n$ e a imagem de T é um subespaço vetorial de $\mathbb{R}^m$.

Com o objetivo de estabelecer a segunda versão do Teorema do Posto, veja o seguinte exemplo.

Exemplo 5.29
Suponha que $T : \mathbb{R}^3 \to \mathbb{R}^4$ é a uma aplicação linear definida por

$$\begin{bmatrix} z \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} a_{11}x + a_{12}y + a_{13}z \\ a_{21}x + a_{22}y + a_{23}z \\ a_{31}x + a_{32}y + a_{33}z \\ a_{41}x + a_{42}y + a_{43}z \end{bmatrix}.$$

Então se $\mathbf{e}_1 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^t$, $\mathbf{e}_2 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^t$ e $\mathbf{e}_3 = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^t$ é a base canônica de $\mathbb{R}^3$, portanto a matriz associada a T é

$$A = \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & T(\mathbf{e}_3) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix}.$$

A imagem de T é exatamente o espaço gerado pela coluna da matriz A e, portanto, a dimensão da imagem de T é igual a dimensão do espaço-coluna e esse, por sua vez, é dado pelo número de colunas-pivô da matriz A. Por outro lado, o núcleo de T consiste em todos os vetores $\mathbf{u} \in \mathbb{R}^3$ para os quais $T(\mathbf{u}) = \mathbf{0}$ e, em termos da matriz A, é exatamente a solução do sistema homogêneo $A\mathbf{x} = \mathbf{0}$, que é a dimensão do núcleo A, a qual é dada pelo número de variáveis livres do sistema $A\mathbf{x} = \mathbf{0}$.

E podemos sintetizar isso por dizer que

$$\begin{aligned} \mathcal{N}(T) &= \mathrm{Nuc}(A), \\ \mathcal{IM}(T) &= \mathrm{Col}(A). \end{aligned}$$

Daí temos que:

Teorema 5.30 (Teorema do Posto, versão 2)
Seja $T : \mathbb{R}^n \to \mathbb{R}^m$ uma aplicação linear, então:

$$\dim \mathcal{N}(T) + \dim \mathcal{IM}(T) = \dim \mathbb{R}^n = n. \tag{5.4}$$

Exemplo 5.31
Seja $T : \mathbb{R}^4 \to \mathbb{R}^3$ a transformação linear definida por

$$\begin{bmatrix} z \\ y \\ z \\ t \end{bmatrix} \mapsto \begin{bmatrix} x - y + z + t \\ 2x - 2y + 3z + 4t \\ 3x - 3y + 4z + 5t \end{bmatrix}.$$

(a) Encontre uma base e a dimensão do $\mathcal{N}(T)$.

Queremos encontrar os vetores $\mathbf{u} = \begin{bmatrix} z & y & z & t \end{bmatrix}^t$, tais que $T(\mathbf{u}) = \mathbf{0}$. Mas isso é equivalente a resolver ao sistema:

$$\begin{cases} x - y + z + t = 0 \\ 2x - 2y + 3z + 4t = 0 \\ 3x - 3y + 4z + 5t = 0 \end{cases}.$$

Escalonando a matriz:

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

As variáveis livres são y e t. Portanto, $\dim \mathcal{N}(T) = 2$. E fazendo $y = 1$, $t = 0$ obtemos $\begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix}^t$ e fazendo $y = 0$, $t = 1$ obtemos $\begin{bmatrix} 1 & 0 & -2 & 1 \end{bmatrix}^t$. Assim,

$$\mathcal{N}(T) = \mathrm{Span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \end{bmatrix} \right\}.$$

(b) Encontre uma base e a dimensão da imagem de T.

Observe que no sistema escalonado anteriormente somente a primeira e a terceira coluna são colunas-pivô. Portanto, o vetor obtido da primeira e da terceira coluna da matriz inicial geram a imagem, isto é,

$$\mathcal{IM}(T) = \operatorname{Span}\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} \right\}.$$

Vamos exibir outra maneira de obter uma base para a $\mathcal{IM}(T)$. Inicialmente, calcule a imagem dos vetores da base canônica do $\mathbb{R}^4$:

$$T\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad T\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ -2 \\ -3 \end{bmatrix}, \quad T\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix}, \quad T\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix}.$$

Sabemos que estes vetores geram a $\mathcal{IM}(T)$. Como $\mathcal{IM}(T)$ é um subespaço vetorial de $\mathbb{R}^3$, segue que qualquer combinação linear desses vetores ainda são vetores da $\mathcal{IM}(T)$, em particular, se aplicarmos as 3 operações elementares. Para isso, monte uma matriz com estes vetores nas linhas e escalone a matriz segundo linhas,

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

Os vetores $\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}^t, \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}^t$ são claramente LI e, também, geram a imagem. Portanto, formam uma base da $\mathcal{IM}(T)$ e novamente temos que a $\dim \mathcal{IM}(T) = 2$.

Assim, podemos confirmar o teorema 5.30 pois temos que

$$\dim \mathcal{N}(T) + \dim \mathcal{IM}(T) = 2 + 2 = 4 = \dim \mathbb{R}^4.$$

Exercícios resolvidos

R5.1. Seja $T : \mathbb{R}^2 \to \mathbb{R}^2$ a aplicação definida por $\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} x + y \\ y \end{bmatrix}$. Mostre que T é linear.

Solução: Precisamos mostrar que T é uma aplicação linear, isto é, que $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ para todo $u, v \in \mathbb{R}^2$ e $T(\alpha\mathbf{u}) = \alpha T(\mathbf{u})$ para todo $\alpha \in \mathbb{R}^2$

e $\mathbf{u} \in \mathbb{R}^2$. Suponha que $\mathbf{u} = \left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$ e $\mathbf{v} = \left[\begin{smallmatrix} x' \\ y' \end{smallmatrix}\right]$, então:

$$\begin{aligned} T(\mathbf{u}+\mathbf{v}) &= T\left(\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} x' \\ y' \end{bmatrix}\right) = T\left(\begin{bmatrix} x+x' \\ y+y' \end{bmatrix}\right) \\ &= \begin{bmatrix} x+x'+y+y' \\ y+y' \end{bmatrix} = \begin{bmatrix} x+y \\ y' \end{bmatrix} + \begin{bmatrix} x'+y' \\ y' \end{bmatrix} = T(\mathbf{u}) + T(\mathbf{v}), \\ T(\alpha\mathbf{u}) &= T\left(\begin{bmatrix} \alpha x \\ \alpha y \end{bmatrix}\right) = \begin{bmatrix} \alpha(x+y) \\ \alpha y \end{bmatrix} = \alpha \begin{bmatrix} x+y \\ y \end{bmatrix} = \alpha T(\mathbf{u}). \end{aligned}$$

E isso demonstra que T é uma aplicação linear. □

R5.2. Seja $T : \mathbb{R}^3 \to \mathbb{R}^2$ a aplicação definida por $\begin{bmatrix} x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} x+y+z \\ 2x-3y+4z \end{bmatrix}$. Mostre que T é linear.

Solução: Considere a matriz:

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & -3 & 4 \end{bmatrix} \text{ e veja que } T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} 1 & 1 & 1 \\ 2 & -3 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

E como vale $A(\mathbf{u}+\mathbf{v}) = A\mathbf{u} + A\mathbf{v}$ e $A(\alpha\mathbf{u}) = \alpha A(\mathbf{u})$, para quaisquer $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$ e $\alpha \in \mathbb{R}$, segue que T é uma aplicação linear. □

R5.3. Mostre que as aplicações abaixo não são lineares.

(a) $T : \mathbb{R}^2 \to \mathbb{R}^2$ a aplicação definida por $\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} xy \\ 2x-3y \end{bmatrix}$;

Solução: Observe que se tomarmos $\mathbf{u} = \left[\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right]$ e $\mathbf{v} = \left[\begin{smallmatrix} 2 \\ 0 \end{smallmatrix}\right]$, então:

$$T(\mathbf{u}+\mathbf{v}) = T\left(\begin{bmatrix} 4 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 4 \\ 5 \end{bmatrix} \text{ e } T(\mathbf{u}) + T(\mathbf{v}) = \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ -3 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \end{bmatrix}.$$

Isso mostra que $T(\mathbf{u}+\mathbf{v})$ não é sempre igual que $T(\mathbf{u}) + T(\mathbf{v})$ e, portanto, T não é uma aplicação linear. □

(b) $T : \mathbb{R}^2 \to \mathbb{R}^3$ a aplicação definida por $\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} x+3 \\ 2x \\ x+y \end{bmatrix}$.

Solução: Observe que

$$T(\mathbf{0}) = T\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Portanto, T não pode ser linear. □

(c) $T:\mathbb{R}^3 \to \mathbb{R}^2$ a aplicação definida por $\begin{bmatrix} x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} |x| \\ 2x-y+z \end{bmatrix}$.

Solução: Para vermos que T não pode ser uma transformação linear, tome $\mathbf{u} = \left[\begin{smallmatrix} -2 \\ 0 \\ -3 \end{smallmatrix}\right]$ e $\mathbf{v} = \left[\begin{smallmatrix} 1 \\ 2 \\ 1 \end{smallmatrix}\right]$, então

$$T(\mathbf{u}+\mathbf{v}) = T\left(\begin{bmatrix} -1 \\ 2 \\ -2 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ -6 \end{bmatrix} \text{ e}$$

$$T(\mathbf{u}) + T(\mathbf{v}) = \begin{bmatrix} 2 \\ -7 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ -6 \end{bmatrix}.$$

Portanto, $T(\mathbf{u}+\mathbf{v}) \neq T(\mathbf{u}) + T(\mathbf{v})$. □

R5.4. Seja $T:\mathbb{R}^2 \to \mathbb{R}^2$ a aplicação, tal que $\begin{bmatrix} 1 \\ 2 \end{bmatrix} \mapsto \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ e $\begin{bmatrix} 0 \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} 1 \\ 4 \end{bmatrix}$. Admita que T é um operador linear e encontre a fórmula para T.

Solução: Vamos começar analisando o que T faz com os vetores da base canônica, para isto, vamos usar que T é linear:

$$\begin{aligned} T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) &= T\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} - 2\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \\ &= T\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right) - 2T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 3 \end{bmatrix} - 2\begin{bmatrix} 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ -5 \end{bmatrix}. \end{aligned}$$

Por outro lado, sabemos que qualquer vetor $\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] = x\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] + y\left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$ e daí,

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = xT\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) + yT\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = x\begin{bmatrix} 2 \\ 3 \end{bmatrix} + y\begin{bmatrix} 0 \\ -5 \end{bmatrix} = \begin{bmatrix} 2x \\ 3x-5y \end{bmatrix}.$$

□

R5.5. Seja $F:\mathbb{R}^2 \to \mathbb{R}^2$ definida por $\begin{bmatrix} x & y & z & t \end{bmatrix}^t \mapsto \begin{bmatrix} x-y+z+t \\ x+2z-t \\ x+y+3z-3t \end{bmatrix}$. Encontre uma base e a dimensão de: (a) da imagem de F, (b) do núcleo de F.

Solução: Vamos começar obtendo o $\mathcal{N}(F)$. Seja A matriz na base canônica de F. Calcular o $\mathcal{N}(F)$ é equivalente a calcular os vetores que satisfazem $A\mathbf{x} = \mathbf{0}$. Escalonando a matriz deste sistema homogêneo obtemos:

$$[A \quad \mathbf{0}] \to \begin{bmatrix} 1 & -1 & 1 & 1 & 0 \\ 1 & 0 & 2 & -1 & 0 \\ 1 & 1 & 3 & -3 & 0 \end{bmatrix} \to$$

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

as soluções do sistema são $x = 2z + t$ e $y = -z + 2t$, portanto, $\mathcal{N}(F) = \text{Span}\left\{\begin{bmatrix} 2 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}\right\}$. Para o cálculo da $\mathcal{IM}(A)$ se lembrarmos que a 1ª e 2ª coluna são colunas pivô da matriz escalonada, então a 1ª e a 2ª colunas de A geram a imagem de A. Segue que $\mathcal{IM}(F) = \text{Span}\left\{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}\right\}$. □

R5.6. Ache a transformação linear a qual é uma rotação anti-horária de 45°.

Solução: Lembre-se de que a matriz associada a uma rotação anti-horária é dada por:

$$A = \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix}.$$

Como o ângulo é de 45° segue que $\theta = \pi/4$, então

$$R\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \sqrt{2}(x-y)/2 \\ \sqrt{2}(x+y)/2 \end{bmatrix}.$$

□

R5.7. Sejam:

$$A = \begin{bmatrix} 1 & 0 & -2 \\ 2 & -1 & -1 \\ 0 & -1 & 1 \end{bmatrix} \text{ e } B = \begin{bmatrix} -3 & 1 & -1 \\ 0 & -1 & 1 \\ 0 & -2 & 0 \end{bmatrix}.$$

Sejam T e S as transformações lineares dadas por A, B, respectivamente. Encontre $R_1 = T \circ S$ e $R_2 = S \circ T$.

Solução: Como T e S são as transformações lineares definidas pelas matrizes A e B, respectivamente, segue que

$$T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x - 2z \\ 2x - y - z \\ -y + z \end{bmatrix} \text{ e } S\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} -3x + y - z \\ -y + z \\ -2y \end{bmatrix}.$$

e daí

$$R_1\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = T\left(S\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right)\right) = \begin{bmatrix} -3x + 5y - z \\ -6x + 5y - 3z \\ -y - z \end{bmatrix}$$

e

$$R_2\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = S\left(T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right)\right) = \begin{bmatrix} -x + 4z \\ -2x + 2z \\ -4x + 2y + 2z \end{bmatrix}.$$

Observe ainda que $R_1 \neq R_2$. □

Exercícios propostos

P5.1. Seja $T : \mathbb{R}^n \to \mathbb{R}^m$ uma transformação linear. Mostre que $\mathcal{N}(T)$ é um subespaço vetorial de $\mathbb{R}^n$ e que $\mathcal{IM}(T)$ é um subespaço vetorial de $\mathbb{R}^m$.

P5.2. Mostre que são lineares as seguintes aplicações: (a) $T : \mathbb{R}^3 \to \mathbb{R}^2$, a aplicação definida por $\begin{bmatrix} x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} x + 2y - 3z \\ 4x - 3y + 6z \end{bmatrix}$; (b) $T : \mathbb{R}^2 \to \mathbb{R}^2$, a aplicação definida por $\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}$, com a, b, c, e d escalares.

P5.3. Seja $F : \mathbb{R}^3 \to \mathbb{R}^3$ uma transformação linear. Se soubermos que:

$$F\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}, \; F\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 5 \\ 2 \\ 7 \end{bmatrix} \text{ e } F\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} -2 \\ 0 \\ 7 \end{bmatrix}.$$

Determine $F(\mathbf{u})$, sendo $\mathbf{u}$ um vetor genérico do $\mathbb{R}^3$.

P5.4. Sejam $F : \mathbb{R}^3 \to \mathbb{R}^2$ e $G : \mathbb{R}^2 \to \mathbb{R}^2$ definidas por $F\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} 2x \\ y + z \end{bmatrix}$ e $G\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} y \\ x \end{bmatrix}$. Calcule (se possível):

a) $F + 2G$,

b) $G \circ F$ e, por fim,

c) $F \circ G$.

P5.5. Admita que os operadores a seguir possuem inversas e encontre a fórmula para T^{-1} quando:

$$\text{(a)} \quad \begin{bmatrix} x \\ y \end{bmatrix} \overset{T}{\mapsto} \begin{bmatrix} x + 2y \\ -2x + 3y \end{bmatrix} \quad \text{e} \quad \text{(b)} \quad \begin{bmatrix} x \\ y \end{bmatrix} \overset{T}{\mapsto} \begin{bmatrix} 2x - 3y \\ x - 4y \end{bmatrix}.$$

P5.6. Para cada uma das matrizes abaixo, determine a aplicação linear associada. Depois encontre uma base e a dimensão do núcleo e da imagem de cada uma delas.

$$(a) \quad \begin{bmatrix} 1 & 2 & 0 & 2 \\ 2 & -1 & 2 & -1 \\ 1 & -3 & 2 & -2 \end{bmatrix}, \qquad (b) \quad \begin{bmatrix} 1 & 0 & 2 & -2 \\ 2 & 3 & 2 & -1 \\ -2 & -3 & 0 & -3 \end{bmatrix}$$

P5.7. Encontre todas as possíveis inversas à direita da matriz:

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

P5.8. Suponha que $T:\mathbb{R}^2\to\mathbb{R}^2$ definido pela matriz $A=\begin{bmatrix}0&1\\1&0\end{bmatrix}$. Encontre (se existirem) vetores $\mathbf{u},\mathbf{v}$, tais que:

a) $T(\mathbf{u})=\mathbf{u}$ e b) $T(\mathbf{v})=-\mathbf{v}$.

P5.9. Ache a transformação linear que é uma rotação anti-horária de 60°.

P5.10. Seja $T:\mathbb{R}^3\to\mathbb{R}^3$ definida por

$$\begin{bmatrix}x\\y\\z\end{bmatrix}\mapsto\begin{bmatrix}x-2y\\z\\x+y\end{bmatrix}.$$

Mostre que T é uma aplicação linear, injetora e sobrejetora, portanto, invertível. Encontre T^{-1}.

P5.11. Sejam $T:\mathbb{R}^n\to\mathbb{R}^m$ uma aplicação linear injetora e se $\{\mathbf{u}_1,\mathbf{u}_2,\dots,\mathbf{u}_k\}$ um conjunto LI. Mostre que:

$$\{T(\mathbf{u}_1),T(\mathbf{u}_2),\dots,T(\mathbf{u}_k)\}\ \text{ também é LI.}$$

P5.12. Seja $T:\mathbb{R}^n\to\mathbb{R}^m$ uma transformação linear com a seguinte propriedade: Se $\{\mathbf{u}_1,\mathbf{u}_2,\dots,\mathbf{u}_k\}$ é uma base do $\mathbb{R}^n$, então $\{T(\mathbf{u}_1),T(\mathbf{u}_2),\dots,T(\mathbf{u}_k)\}$ é linearmente independente em $\mathbb{R}^m$. Prove então que T é injetora.

Capítulo 6

Determinantes

Já em 1683, o japonês Seki Takakazu inventou o conceito de determinante, provavelmente reinventado em 1693 pelo alemão Gottfried Liebniz, pois não havia comunicação entre eles. Na época, o determinante estava relacionado com as fórmulas para exprimir a solução de um sistema linear de n equações e n variáveis, uma vez que a teoria de matrizes só seria desenvolvida muito mais tarde. Posteriormente, em 1812, Augustin-Louis Cauchy identificou que o determinante poderia ser usado para calcular a área do paralelogramo ou o volume do paralelepípedo. Somente depois o determinante seria associado com as formas multilineares alternadas.

O determinante associa um número para cada matriz quadrada.

O principal uso do determinante está no fato de que o determinante de um operador linear é não-nulo se, e somente se, o operador é invertível.

Neste capítulo iremos deduzir fórmulas e procedimentos para calcular o determinante de uma matriz, além de apresentar dois métodos para obter a inversa de uma matriz e um critério para decidir se um sistema admite solução única.

6.1 Determinantes de ordens $1, 2$ e 3

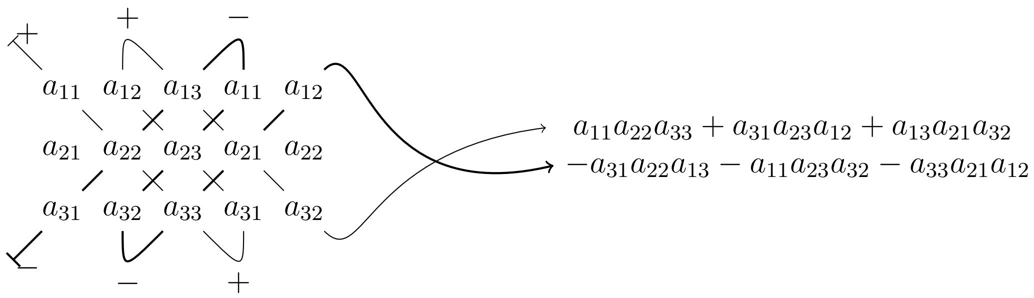
O determinante é uma função que toma uma matriz quadrada A e retorna com um número.

Os determinantes de ordens $1, 2$ e 3 são definidos por:

$$\det \begin{bmatrix} a_{11} \end{bmatrix} = a_{11}, \quad \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{21}a_{12} \text{ e}$$

$$\begin{aligned} \det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} &= a_{11}a_{22}a_{33} + a_{13}a_{21}a_{32} + a_{23}a_{31}a_{12} \\ &\quad - a_{31}a_{22}a_{13} - a_{32}a_{23}a_{11} - a_{33}a_{21}a_{12}. \end{aligned}$$

Observe que o determinante da matriz 3×3 possui seis produtos, cada um consiste de três elementos da matriz original. Três destes produtos recebem sinal positivo e três deles recebem sinal negativo. O diagrama a seguir ajuda a memorizar a fórmula envolvida no cálculo do determinante, o esquema é obtido por repetir a 1ª e 2ª coluna no final da matriz. O determinante é obtido através da soma dos produtos, ao longo das três setas assinaladas com o sinal $+$, mais a soma dos negativos dos produtos dos elementos que estão nas setas assinaladas com o sinal $-$.



Exemplo 6.1
Sejam $A = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 5 & -2 \\ 2 & -3 & 4 \end{bmatrix}$ e $B = \begin{bmatrix} 3 & 2 & 0 \\ -4 & 0 & -1 \\ 1 & -3 & 2 \end{bmatrix}$. Encontre $\det(A)$ e $\det(B)$.

$$\begin{aligned} \det(A) &= 2(5)4 + 1(-2)2 + 1(0)(-3) - 2(5)1 - (-3)(-2)2 - 4(0)1 \\ &= 40 - 4 - 10 - 12 = 14, \\ \det(B) &= 0 + (-2) + 0 - 0 - 9 - (-16) \\ &= 5. \end{aligned}$$

6.2 Determinante em Geral

Observe que o determinante de uma matriz $A = [a_{ij}]$ de ordem 3 pode ser reescrito da seguinte forma:

$$\begin{aligned} \det(A) &= a_{11}a_{22}a_{33} + a_{13}a_{21}a_{32} + a_{23}a_{31}a_{12} - a_{31}a_{22}a_{13} - a_{32}a_{23}a_{11} - a_{33}a_{21}a_{12} \\ &= a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} - a_{12}a_{33}a_{21} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{31}a_{22} \\ &= a_{11}(a_{22}a_{33} - a_{32}a_{23}) - a_{12}(a_{21}a_{33} - a_{31}a_{23}) + a_{13}(a_{21}a_{32} - a_{31}a_{22}) \\ &= a_{11} \det \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} - a_{12} \det \begin{bmatrix} a_{21} & a_{23} \\ a_{13} & a_{33} \end{bmatrix} + a_{13} \det \begin{bmatrix} a_{21} & a_{31} \\ a_{22} & a_{32} \end{bmatrix}. \end{aligned}$$

Algo parecido pode ser feito com a matriz A de ordem 2:

$$\det(A) = a_{11} \det [a_{22}] - a_{12} \det [a_{21}] .$$

Para facilitar a expressão do determinante introduzimos a notação a seguir.

Definição 6.2
Sejam $n > 1$ e A uma matriz quadrada de ordem n, indicaremos por A_{ij} a matriz de ordem $n-1$, obtida de A por apagar a i-ésima linha e a j-ésima coluna e por $\Delta_{ij} = (-1)^{i+j} \det(A_{ij})$. Chamamos Δ_{ij} de **cofator** de A, na linha i e coluna j.

Exemplo 6.3

Considere $A = \begin{bmatrix} 1 & -1 & 2 \\ -1 & 1 & 0 \\ 2 & -3 & 1 \end{bmatrix}$. Então,

$$A_{11} = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}, \; A_{23} = \begin{bmatrix} 1 & -1 \\ 2 & -3 \end{bmatrix} \text{ e } A_{21} = \begin{bmatrix} -1 & 2 \\ -3 & 1 \end{bmatrix}.$$

E os cofatores correspondentes são:

$\Delta_{11} = (-1)^{1+1}\det(A_{11}) = 1$, $\Delta_{23} = (-1)^{2+3}\det(A_{23}) = -(-3+2) = 1$ e
$\Delta_{21} = (-1)^{2+1}\det(A_{21}) = -(-1+6) = -5$.

Com essa notação podemos reescrever o determinante no caso de A ser de ordem 2

$$\det(A) = a_{11}\Delta_{22} + a_{12}\Delta_{21}.$$

e no caso de A ser uma matriz de ordem 3, o determinante fica:

$$\det(A) = a_{11}\Delta_{11} + a_{12}\Delta_{12} + a_{13}\Delta_{13}.$$

Recursivamente definimos para a matriz 4×4 o seu determinante, por ser

$$\det(A) = a_{11}\Delta_{11} + a_{12}\Delta_{12} + a_{13}\Delta_{13} + a_{14}\Delta_{14}.$$

E assim, sucessivamente, isto é,

Definição 6.4

Seja $n > 1$ e A uma matriz quadrada de ordem n, definimos o determinante de A como sendo:

$$\det(A) = a_{11}\Delta_{11} + a_{12}\Delta_{12} + \cdots + a_{1n}\Delta_{1n}.$$

Exemplo 6.5

Considere a matriz $A = \begin{bmatrix} 0 & 1 & 1 & -1 \\ 3 & 0 & 1 & 0 \\ 1 & -1 & 0 & -2 \\ 2 & -1 & -1 & 1 \end{bmatrix}$, calcule $\det(A)$.

$$\begin{aligned} \det(A) &= 0\Delta_{11} + 1\Delta_{12} + 1\Delta_{13} + (-1)\Delta_{14} \\ &= 0(3) + 1(11) + 1(-9) + (-1)(-4) = 6. \end{aligned}$$

Apesar de termos definido o determinante para qualquer matriz quadrada de ordem n, sempre que precisamos calcular o determinante de uma matriz de ordem n precisamos calcular n determinantes de matrizes de ordem $n-1$. Quando n cresce, a quantidade de operações necessárias para calcular o determinante cresce muito depressa, o que inviabiliza a operação, para perceber isso tente ver quantas operações você necessitaria se tivesse que calcular o determinante de uma matriz de ordem $n = 7$. Mesmo empregando um computador para fazer esta tarefa, se utilizarmos essa fórmula, mesmo para matrizes relativamente pequenas, por exemplo $n = 15$, qualquer computador terá dificuldades em retornar uma resposta.

Na atualidade os softwares empregados no cálculo do determinante usam outros métodos. Para entendermos como esses métodos funcionam vamos observar algumas propriedades do determinante. Para isto veja a definição do seguinte conceito:

Definição 6.6
Dizemos que uma função $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ é 2-linear, se ela for linear em cada uma de suas entradas, isto é, para cada x, y, z e $\alpha \in \mathbb{R}$ valem:

$$f(x+y, z) = f(x,z) + f(y,z), \quad f(x, y+z) = f(x,y) + f(x,z)$$
$$f(\alpha x, y) = \alpha f(x,y) \text{ e } f(x, \alpha y) = \alpha f(x,y).$$

Exemplo 6.7
Considere $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ definida por $f(x,y) = xy$, então se $f(x+x', y) = (x+x')y = xy + x'y = f(x,y) + f(x', y)$ e se $f(x, y+y') = x(y+y') = xy + xy' = f(x,y) + f(x,y')$. Além disso, se multiplicarmos x ou y por um número α temos:

$$f(\alpha x, y) = (\alpha x)y = \alpha(xy) = \alpha f(x,y) \text{ e } f(x, \alpha y) = x(\alpha y) = \alpha(xy) = \alpha f(x,y).$$

Considere A uma matriz $n \times n$. Podemos pensar tal matriz como n colunas $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n$, onde cada uma destas colunas é um vetor $\mathbb{R}^n$, e podemos escrever $A = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \end{bmatrix}$. Vamos definir a função D que toma n vetores do $\mathbb{R}^n$ e retorna um número por

$$D(\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n) = \det \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \end{bmatrix}.$$

Observe que a função determinante é uma função que toma uma matriz quadrada e retorna um número.

Exemplo 6.8
Calcule $D\left(\begin{bmatrix} 1 \\ -3 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \end{bmatrix}\right)$. Pela definição da função D temos:

$$D\left(\begin{bmatrix} 1 \\ -3 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \end{bmatrix}\right) = \det \begin{bmatrix} 1 & 2 \\ -3 & -1 \end{bmatrix} = -1 + 6 = 5.$$

Proposição 6.9
A função $D(\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n)$ satisfaz as seguintes propriedades:

(d_1) *D é **alternada**, isto é, se $\mathbf{c}_i = \mathbf{c}_j$ para $i \neq j$ então $D(\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n) = 0$.*

(d_2) *D é n-linear, isto é, D é linear em cada uma de suas colunas. Mais precisamente, se todos os $\mathbf{c}_j$ com $j \neq i$ estiverem fixos, então*

$$D(\mathbf{c}_1, \ldots, \mathbf{c}_i + \lambda \mathbf{c}'_i, \ldots, \mathbf{c}_n) = D(\mathbf{c}_1, \ldots, \mathbf{c}_i, \ldots, \mathbf{c}_n) + \lambda D(\mathbf{c}_1, \ldots, \mathbf{c}'_i, \ldots, \mathbf{c}_n).$$

(d_3) *Se $[\mathbf{e}_1, \ldots, \mathbf{e}_n]$ é a matriz identidade então $D(\mathbf{e}_1, \ldots, \mathbf{e}_n) = 1$.*

A principal implicação das propriedades acima está na próxima observação.

Observação 6.10
Considere a função D, como a que satisfaz as condições $(d_1) - (d_3)$. Então, a função é antissimétrica, isto é, se trocarmos $\mathbf{c}_i$ por $\mathbf{c}_j$ o valor de D troca de sinal. Mais precisamente,

$$D(\mathbf{c}_1, \ldots, \mathbf{c}_i, \ldots, \mathbf{c}_j, \ldots, \mathbf{c}_n) = -D(\mathbf{c}_1, \ldots, \mathbf{c}_j, \ldots, \mathbf{c}_i, \ldots, \mathbf{c}_n).$$

Vamos demonstrar esse fato. Para simplificar a notação e, como só vamos tratar dos vetores $\mathbf{c}_i$ e $\mathbf{c}_j$ e os outros vetores irão permanecer fixos, considere $D(\mathbf{c}_i, \mathbf{c}_j) = D(\mathbf{c}_1, \ldots, \mathbf{c}_i, \ldots, \mathbf{c}_j, \ldots, \mathbf{c}_n)$. Temos:

$$\begin{aligned}
0 = D(\mathbf{c}_i + \mathbf{c}_j, \mathbf{c}_i + \mathbf{c}_j) &= D(\mathbf{c}_i, \mathbf{c}_i + \mathbf{c}_j) + D(\mathbf{c}_j, \mathbf{c}_i + \mathbf{c}_j) \\
&= D(\mathbf{c}_i, \mathbf{c}_i) + D(\mathbf{c}_i, \mathbf{c}_j) + D(\mathbf{c}_j, \mathbf{c}_i) + D(\mathbf{c}_j, \mathbf{c}_j) \\
&= D(\mathbf{c}_i, \mathbf{c}_j) + D(\mathbf{c}_j, \mathbf{c}_i).
\end{aligned}$$

E daí $D(\mathbf{c}_i, \mathbf{c}_j) = -D(\mathbf{c}_j, \mathbf{c}_i)$. Portanto, se estivermos calculando o determinante de uma matriz e trocarmos duas colunas entre si, o determinante troca de sinal.

Com estas propriedades também conseguimos reobter a fórmula para o determinante. Veja o próximo exemplo.

Exemplo 6.11
Considere a matriz $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, logo a coluna $\mathbf{c}_1 = \begin{bmatrix} a \\ c \end{bmatrix}$ e $\mathbf{c}_2 = \begin{bmatrix} b \\ d \end{bmatrix}$. Observe que $\begin{bmatrix} a \\ c \end{bmatrix} = a \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, logo pela propriedade (d_2) temos que:

$$\begin{aligned}
D\left(\begin{bmatrix} a \\ c \end{bmatrix}, \begin{bmatrix} b \\ d \end{bmatrix}\right) &= D\left(a\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ c \end{bmatrix}, \begin{bmatrix} b \\ d \end{bmatrix}\right) \\
&= aD\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} b \\ d \end{bmatrix}\right) + D\left(\begin{bmatrix} 0 \\ c \end{bmatrix}, \begin{bmatrix} b \\ d \end{bmatrix}\right) \\
&= aD\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} b \\ d \end{bmatrix}\right) + D\left(c\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} b \\ d \end{bmatrix}\right) \\
&= aD\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} b \\ d \end{bmatrix}\right) + cD\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} b \\ d \end{bmatrix}\right) \\
&= aD\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, b\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ d \end{bmatrix}\right) + cD\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}, b\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ d \end{bmatrix}\right) \\
&= abD\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) + adD\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \\
&\qquad + cbD\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) + cdD\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \\
&= adD\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) - cbD\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = ad - bc.
\end{aligned}$$

Observe que usamos a propriedade (d_1) quando trocamos a primeira coluna com a segunda e, por isso, trocamos o sinal e na última igualdade usamos (d_3).

6.3 Matriz de Permutação e o Determinante da Transposta

Definição 6.12
Uma **matriz de permutação** é uma matriz obtida da matriz identidade pela permutação de suas colunas.

Exemplo 6.13
A matriz
$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$
é obtida da matriz identidade por permutar a primeira e a segunda coluna. Além disso, é claro que o $\det(P) = -1$, pois o determinante é igual ao determinante da matriz identidade multiplicado por (-1).

Se $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ é a base canônica do $\mathbb{R}^n$, denotamos a matriz de permutação por $P_{i_1 i_2 \ldots i_n}$, isso significa que na primeira coluna temos o vetor $\mathbf{e}_{i_1}$, na segunda coluna está o vetor $\mathbf{e}_{i_2}$ e na n-ésima coluna está o vetor $\mathbf{e}_{i_n}$, assim, a matriz acima é denotada por P_{213}.

O determinante de uma matriz de permutação é sempre ± 1, uma vez que podemos obter a matriz identidade depois de executarmos um número finito de permutações em suas colunas. Podemos ser mais precisos: se executamos um número par de trocas, então o determinante é 1; e se executamos um número ímpar de trocas, então o determinante é -1. Logo, definimos o sinal da matriz de permutação P como:
$$\sigma(P) = \det(P).$$

Agora vamos enunciar o resultado principal desta seção.

Teorema 6.14
Para todo $n > 1$ se A é uma matriz $n \times n$, então $\det(A) = \det(A^t)$.

Disso seguem duas observações importantes. Para entendermos bem as observações considere:
$$A = [a_{ij}] = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \end{bmatrix} = \begin{bmatrix} \ell_1 \\ \ell_2 \\ \vdots \\ \ell_n \end{bmatrix}$$
escrita como n colunas ou n linhas.

Observação 6.15
Dada uma matriz $A = [a_{ij}]$ de ordem n podemos considerar $B = A^t$, logo o $\det(B) = \det(A)$ e D é n-linear e alternada sobre as colunas de B, mas isso quer dizer que, com respeito a matriz A, $D(A)$ é n-linear e alternada com respeito às linhas de A.

Observação 6.16
Sejam A e $B = A^t$. Calculando o determinante por fazer a expansão em termos da primeira linha de B, temos

$$\begin{aligned}\det A = \det B &= b_{11}\Delta_{11} + b_{12}\Delta_{12} + \cdots + b_{1n}\Delta_{1n}\\ &= a_{11}(-1)^{1+1}\det A^t_{11} + a_{21}(-1)^{2+1}\det A^t_{21} + \cdots + a_{n1}(-1)^{n+1}\det A^t_{n1}\\ &= a_{11}(-1)^{1+1}\det A_{11} + a_{21}(-1)^{2+1}\det A_{21} + \cdots + a_{n1}(-1)^{n+1}\det A_{n1},\end{aligned}$$

isto é, podemos calcular o determinante por fazer a expansão segundo a primeira coluna de A. Na verdade, podemos calcular o determinante por fazer a expansão em qualquer linha e qualquer coluna. Para ver como isso é feito veja o exercício R6.6.

6.4 Regra de Cramer

Sejam $A = [a_{ij}]$ uma matriz $n \times n$ e $\mathbf{b} \in \mathbb{R}^n$ um vetor. Considere a equação

$$A\mathbf{x} = \mathbf{b}.$$

Seja A_k a matriz obtida de A por substituir a coluna k de A pelo vetor $\mathbf{b}$. Então, vale o seguinte resultado:

Teorema 6.17
O sistema (quadrado) $A\mathbf{x} = \mathbf{b}$ possui uma única solução se, e só se, $\det(A) \neq 0$. Nesse caso, a solução é dada por

$$x_1 = \frac{\det(A_1)}{\det(A)}, \quad x_2 = \frac{\det(A_2)}{\det(A)}, \quad \ldots, \quad x_n = \frac{\det(A_n)}{\det(A)}.$$

Veja o seguinte exemplo

Exemplo 6.18
Resolva o seguinte sistema de equações lineares

$$\begin{cases} x + y + z = 5 \\ x - 2y - 3z = -1 \\ 2x + y - z = 3 \end{cases}.$$

Para usarmos o teorema anterior, precisamos calcular o seguinte determinante:

$$\det(A) = \det\begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & -3 \\ 2 & 1 & -1 \end{bmatrix} = 2 - 6 + 1 + 4 + 3 + 1 = 5.$$

Como $\det(A) \neq 0$, o sistema tem apenas uma solução, que é dada por

$$x = \frac{1}{5}\det\begin{bmatrix} 5 & 1 & 1 \\ -1 & -2 & -3 \\ 3 & 1 & -1 \end{bmatrix} = \frac{20}{5}, \quad y = \frac{1}{5}\det\begin{bmatrix} 1 & 5 & 1 \\ 1 & -1 & -3 \\ 2 & 3 & -1 \end{bmatrix} = \frac{-10}{5}$$

$$z = \frac{1}{5}\det\begin{bmatrix} 1 & 1 & 5 \\ 1 & -2 & -1 \\ 2 & 1 & 3 \end{bmatrix} = \frac{15}{5}.$$

Seja $A = [a_{ij}]$ uma matriz $n \times n$ e, como já explicamos, chamamos os elementos Δ_{ij} de cofatores da matriz A na posição ij. A **Adjunta Clássica** de A, denotada por $\operatorname{adj}(A)$, é a transposta da matriz de cofatores de A, a saber:

$$\operatorname{adj}(A) = \begin{bmatrix} \Delta_{11} & \Delta_{21} & \cdots & \Delta_{n1} \\ \Delta_{12} & \Delta_{22} & \cdots & \Delta_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta_{1n} & \Delta_{2n} & \cdots & \Delta_{nn} \end{bmatrix}.$$

Chamamos de "Adjunta Clássica", em vez de simplesmente "Adjunta", porque, hoje em dia, o termo "adjunta" é reservado para outro conceito totalmente diferente.

Teorema 6.19
Seja A uma matriz quadrada de ordem n qualquer. Então,

$$\operatorname{adj}(A)A = \det(A)I$$

sendo I a matriz identidade. Assim, se $\det(A) \neq 0$,

$$A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A).$$

Exemplo 6.20
Por utilizar a técnica sugerida no teorema acima vamos obter a inversa de

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

Para isso, precisamos determinar os cofatores Δ_{ij} da matriz A. Vamos construir uma matriz intermediária D e, por fim, obter a $\operatorname{adj}(A)$ que é D^t.

$$D = \begin{bmatrix} \Delta_{11} & \Delta_{12} & \Delta_{13} \\ \Delta_{21} & \Delta_{22} & \Delta_{23} \\ \Delta_{31} & \Delta_{32} & \Delta_{33} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ -5 & -3 & 1 \\ -2 & -2 & 0 \end{bmatrix}.$$

Portanto,

$$\operatorname{adj}(A) = D^t = \begin{bmatrix} 1 & -5 & -2 \\ 1 & -3 & -2 \\ 1 & 1 & 0 \end{bmatrix}.$$

Sabendo que $\det(A) = 2$ segue que

$$A^{-1} = \frac{1}{\det A} \operatorname{adj}(A) = \begin{bmatrix} 1/2 & -5/2 & -1 \\ 1/2 & -3/2 & -1 \\ 1/2 & 1/2 & 0 \end{bmatrix}.$$

6.5 Determinante do Produto

Antes de obtermos este resultado vamos ver como se comporta o determinante, quando o aplicamos em matrizes elementares:

(a) Se E_1 é a matriz obtida por executar $\ell_i \to \ell_i + k\ell_j$ na matriz identidade, então, $\det(E_1) = 1$;

(b) Se E_2 é a matriz obtida por executar $\ell_i \to k\ell_i$ com $k \neq 0$ na matriz identidade, então, $\det(E_2) = k$;

(c) Se E_3 é a matriz obtida por executar $\ell_i \leftrightarrow \ell_j$ na matriz identidade, então, $\det(E_3) = -1$;

(d) Se A é uma matriz qualquer e E é uma matriz elementar, então, $\det(EA) = \det(E)\det(A)$.

Teorema 6.21
Se A e B são matrizes $n \times n$, então $\det(AB) = \det(A)\det(B)$.

Demonstração: (1ª caso) Se A ou B não são invertíveis, logo pode acontecer: a) A invertível e B não é; b) B invertível e A não é; c) A e B são ambas não invertíveis. Então, nas três situações, podemos concluir que AB também não é invertível. De fato, se a) ocorre então existe A^{-1} e admita que AB seja invertível, nesse caso, $B = A^{-1}(AB)$ também é invertível. Se ocorrer b) tratamos da mesma maneira. Se ocorrer c) suponha, por absurdo, que AB é invertível, nesse caso, existe C, tal que $C(AB) = I = (CA)B$ e, portanto, B é invertível, o que é um absurdo.

Logo, se A ou B não é invertível, então AB também não é invertível e $\det(AB) = 0$ e como $\det(A) = 0$ ou $\det(B) = 0$ segue a igualdade.

(2ª caso) Se A e B são invertíveis, sabemos que existe uma sequência finita de operações elementares (sobre as linhas) $E_1, \ldots, E_k$ que tornam A a matriz identidade, isto é, $I = E_k^{-1} \cdots E_1^{-1} A$ Daí temos:

$$\det(AB) = \det(E_1 \cdots E_k I B)$$

e aplicando um número finito de vezes a propriedade (d) acima obtemos:

$$\begin{aligned}
\det(AB) &= \det(E_1)\det(E_2 \cdots E_{k-1}B) \\
&= \det(E_1)\det(E_2)\cdots\det(E_{k-1}E_k)\det(B) \\
&= \det(E_1)\det(E_2)\cdots\det(E_{k-1})\det(E_k)\det(B) \\
&= \det(E_1)\det(E_2)\cdots\det(E_{k-1}E_k)\det(B) \\
&= \det(E_1E_2\cdots E_{k-1}E_kI)\det(B) = \det(A)\det(B).
\end{aligned}$$

□

Teorema 6.22
Seja A uma matriz $n \times n$. A matriz A é invertível se, e somente se, $\det(A) \neq 0$, *e neste caso* $\det(A^{-1}) = \frac{1}{\det(A)}$.

Demonstração: ($\Rightarrow$) Se A é invertível existe uma matriz B, tal que $AB = I$, aplicando o determinante dos dois lados desta igualdade obtemos:

$$\det(AB) = \det(A)\det(B) = \det(I) = 1.$$

Isso implica que $\det(A) \neq 0$ e também que $\det(A^{-1}) = \frac{1}{\det(A)}$.

($\Leftarrow$) Se $\det(A) \neq 0$ e seja B a adjunta clássica de A, então sabemos que $BA = \det(A)I$, daí temos que $\frac{1}{\det(A)}BA = I$, isto é, ao multiplicarmos $\frac{1}{\det(A)}B$ por A obtemos I, e como a inversa de uma matriz é única, segue que $A^{-1} = \frac{1}{\det(A)}B$. $\square$

6.6 Matrizes em Blocos

O principal resultado dessa seção é o seguinte:

Teorema 6.23
Seja B uma matriz quadrada triangular inferior (superior) em blocos com os blocos diagonais $A_1, A_2, \ldots, A_k$. Então,

$$\det(B) = \det(A_1)\det(A_2)\cdots\det(A_k).$$

Exemplo 6.24
Calcule o determinante de $B = \begin{bmatrix} 3 & 2 & 7 & -9 & -4 \\ 5 & -1 & 1 & -1 & -9 \\ 0 & 0 & 3 & 2 & 0 \\ 0 & 0 & -4 & 0 & -1 \\ 0 & 0 & 1 & -3 & 2 \end{bmatrix}$. Observe que B é uma matriz triangular superior em blocos. Pelo teorema basta calcularmos o determinante de cada bloco diagonal:

$$\det\begin{bmatrix} 3 & 2 \\ 5 & -1 \end{bmatrix} = -10 - 3 = -13, \qquad \det\begin{bmatrix} 3 & 2 & 0 \\ -4 & 0 & -1 \\ 1 & -3 & 2 \end{bmatrix} = 5.$$

Então, $\det(B) = (-13)(5) = -65$.

Corolário 6.25
Seja $B = [b_{ij}]$ uma matriz quadrada triangular inferior (superior). Então,

$$\det(B) = b_{11}b_{22}\cdots b_{nn}.$$

Observação 6.26
Seja $N = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$, em que A, B, C e D são matrizes quadradas. Em geral, não é válido que $\det(N) = \det(A)\det(D) - \det(B)\det(C)$. (confira o exercício P6.8).

6.7 Área e Volume através do Determinante

Nesta seção mostraremos que os determinantes podem ser usados para calcular a área e o volume, tal como foi mencionado na introdução do capítulo. Faremos apenas para o $\mathbb{R}^2$ e $\mathbb{R}^3$. E generalizamos o conceito de volume, usando estes métodos para espaços de dimensão n maior que 3.

Teorema 6.27
Se A é uma matriz de ordem 2, a área do paralelogramo determinado pelas colunas de A é igual ao $|\det(A)|$. Se A é de ordem 3, o volume do paralelepípedo determinado pelas colunas de A é igual ao $|\det(A)|$.

Demonstração: No caso de A ser de ordem 2, o resultado é verdadeiro se A for diagonal, pois

$$\left|\det \begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix}\right| = |ad| = \{\text{área do retângulo de lados a e d}\}.$$

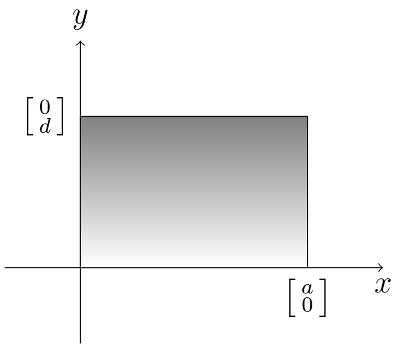


Figura 6.1: Área= $|ad|$

Suponha que $A = [\mathbf{c}_1 \quad \mathbf{c}_2]$ é uma matriz qualquer. Para provarmos o resultado basta verificarmos que a matriz A pode ser transformada em uma matriz diagonal sem que com isso altere o $|\det(A)|$ e nem a área do paralelogramo. Já sabemos que trocar uma coluna com a outra não altera o valor de $|\det(A)|$, assim como somar a uma coluna um múltiplo da outra coluna (verifique que isso é o suficiente para transformar qualquer matriz em uma matriz diagonal).

Claramente a área do paralelogramo com respeito aos vetores $\mathbf{c}_1, \mathbf{c}_2$ é a mesma que com respeito a $\mathbf{c}_2, \mathbf{c}_1$. Além disso, se chamarmos a reta determinada por $\mathbf{0}$ e $\mathbf{c}_1$ de L, então a reta $\mathbf{c}_2 + L$ é uma reta paralela a L, e $\mathbf{c}_2 + t\mathbf{c}_1$ pertence a reta $\mathbf{c}_2 + L$ para todo t. Como a área de um paralelogramo é o comprimento da base vezes a sua altura com respeito a esta base, segue que a área do paralelogramo determinado por $\mathbf{c}_1, \mathbf{c}_2$ é sempre igual a área do paralelogramo determinado por $\mathbf{c}_1, \mathbf{c}_2 + t\mathbf{c}_1$. Veja a próxima figura para entender melhor o que acontece.

No caso de A ter ordem 3, o raciocínio é semelhante. No caso em que A é diagonal o resultado é claramente verdadeiro.

Se $A = [\mathbf{c}_1 \quad \mathbf{c}_2 \quad \mathbf{c}_3]$ é uma matriz qualquer, podemos transformá-la em uma matriz diagonal, por permutar as suas colunas e somar a uma coluna um múltiplo

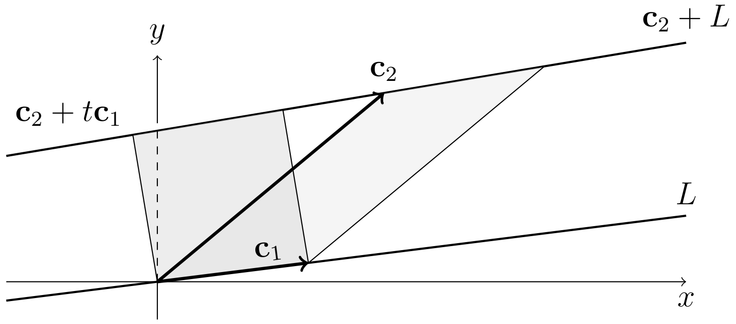


Figura 6.2: Área$= |ad - bc|$

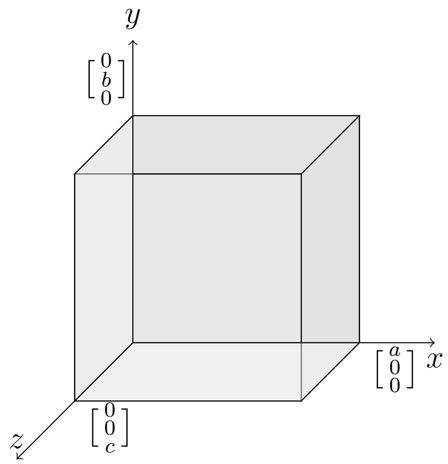


Figura 6.3: Volume do paralelepípedo é $|abc|$

de outra. Claramente estas operações não alteram o $|\det(A)|$. Vamos ver que essas operações não alteram o volume do paralelepípedo determinado pelos vetores $\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3$. Lembre-se de que o volume de um paralelepípedo é determinado pela multiplicação da área de uma de suas faces pela altura com respeito a essa face. Vamos considerar a face determinada pelos vetores $\mathbf{c}_1, \mathbf{c}_3$. Observe que, pelo mesmo argumento usado no $\mathbb{R}^2$, a área dessa face não se altera se trocarmos os vetores $\mathbf{c}_1, \mathbf{c}_2$ por $\mathbf{c}_1, \mathbf{c}_2 + t\mathbf{c}_1$ qualquer que seja $t \in \mathbb{R}$.

Por simplicidade vamos supor que a face determinada por $\mathbf{c}_1, \mathbf{c}_3$ coincida com o plano xz veja a próxima figura

Considere o plano $P = \operatorname{Span}\{\mathbf{c}_1, \mathbf{c}_3\}$, então a face de cima do paralelepípedo está no plano $P+\mathbf{c}_2$ obtido por transladar o plano P. O volume do paralelepípedo é igual a área determinada por $\mathbf{c}_1, \mathbf{c}_3$, vezes a altura de $\mathbf{c}_2$ com respeito ao plano P. Todos vetores da forma $\mathbf{c}_2 + r\mathbf{c}_1$ e $\mathbf{c}_2 + t\mathbf{c}_3$, com $r, t \in \mathbb{R}$, tem a mesma altura com respeito ao plano P, uma vez que se encontram no plano $P + \mathbf{c}_2$ que é paralelo a P. Portanto, o volume do paralelepípedo não se altera quando trocamos $[\mathbf{c}_1 \quad \mathbf{c}_2 \quad \mathbf{c}_3]$ por $[\mathbf{c}_1 \quad \mathbf{c}_2 + r\mathbf{c}_1 \quad \mathbf{c}_3]$, uma vez que isso é equivalente a deslizar a face superior do paralelepípedo no plano $P + \mathbf{c}_2$. Veja a figura abaixo

$\square$

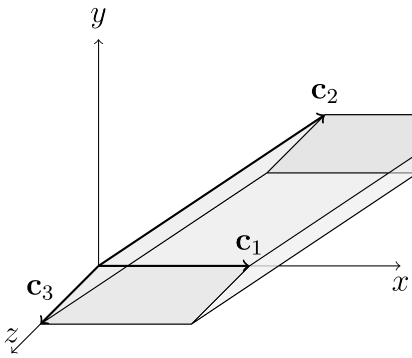


Figura 6.4: Paralelepípedo

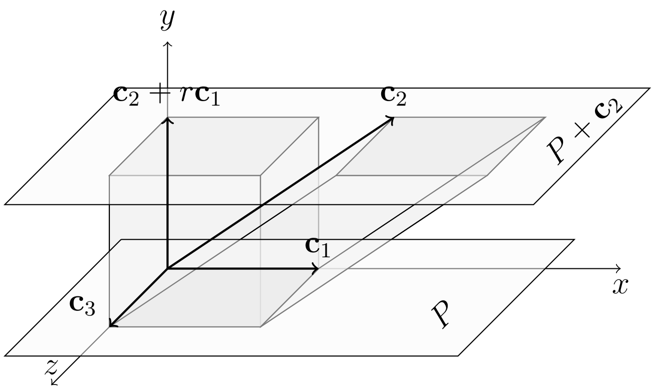


Figura 6.5: Deslizando a face do paralelepípedo

Exemplo 6.28
Calcule a área do paralelogramo determinado pelos pontos $\left[\begin{smallmatrix}-2\\-2\end{smallmatrix}\right]$, $\left[\begin{smallmatrix}0\\3\end{smallmatrix}\right]$, $\left[\begin{smallmatrix}4\\-1\end{smallmatrix}\right]$ e $\left[\begin{smallmatrix}6\\4\end{smallmatrix}\right]$.

Para começar, translade o paralelogramo até que o vértice $\left[\begin{smallmatrix}-2\\-2\end{smallmatrix}\right]$ coincida com a origem. Para isso, subtraia $\left[\begin{smallmatrix}-2\\-2\end{smallmatrix}\right]$ de todos os vértices. O novo paralelogramo tem a mesma área e vértices $\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$, $\left[\begin{smallmatrix}2\\5\end{smallmatrix}\right]$, $\left[\begin{smallmatrix}6\\1\end{smallmatrix}\right]$ e $\left[\begin{smallmatrix}8\\6\end{smallmatrix}\right]$. Logo, o paralelogramo é determinado pelas colunas de

$$A = \begin{bmatrix} 2 & 6 \\ 5 & 1 \end{bmatrix},$$

e daí, $|\det(A)| = |-28| = 28\,uni^2$ é a área do paralelogramo inicial.

Exercícios resolvidos

R6.1. Calcule o determinante de cada uma das matrizes seguintes.

$$a)\quad A = \begin{bmatrix} 6 & 5 \\ 2 & 3 \end{bmatrix}, \qquad b)\quad B = \begin{bmatrix} 2 & -3 \\ 4 & 7 \end{bmatrix}, \qquad c)\quad C = \begin{bmatrix} t-5 & 6 \\ 3 & t+2 \end{bmatrix}.$$

Solução: Usando a fórmula do determinante 2×2 temos:

a) $\det(A) = 6(3) - 5(2) = 18 - 10 = 8$,

b) $\det(B) = 14 + 12 = 26$,

c) $\det(C) = (t-5)(t+2) - 18 = t^2 - 3t - 10 - 18 = t^2 - 10t - 28$. □

R6.2. Calcule o determinante de cada uma das matrizes seguintes.

$$a)\quad A = \begin{bmatrix} 2 & 3 & 4 \\ -1 & 2 & 3 \\ -4 & 0 & 3 \end{bmatrix}, \qquad b)\quad B = \begin{bmatrix} 2 & 1 & -5 \\ -2 & 3 & 2 \\ -2 & 1 & 1 \end{bmatrix}.$$

Solução: Usando a fórmula do determinante 3×3 e escalonando temos:

a) $\det(A) = 12 - 36 + 0 - (-32) - 0 - (-9) = 17$,

b) fazendo $\ell_2 \to \ell_2 + \ell_1$ e $\ell_3 \to \ell_3 + \ell_1$ obtemos que

$$\det(B) = \det \begin{bmatrix} 2 & 1 & -5 \\ 0 & 4 & -3 \\ 0 & 2 & -4 \end{bmatrix} = 2(-16 - (-6)) = -20.$$ □

R6.3. Calcule o determinante de cada uma das matrizes seguintes:

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

Solução: Em a), se fizermos $\ell_2 \to \ell_2 + 2\ell_1$, $\ell_3 \to \ell_3 - \ell_1$, $\ell_4 \to \ell_4 + \ell_1$, $\ell_3 \longleftrightarrow \ell_4$ e, por fim, $\ell_3 \to \ell_3 + 4\ell_1$ ficamos com a matriz:

$$\det \begin{bmatrix} 1 & 2 & -3 & -2 \\ 0 & 1 & -4 & -9 \\ 0 & 0 & -23 & -35 \\ 0 & 0 & 0 & 4 \end{bmatrix} = (-1)(1)(1)(-23)(4) = 92.$$

Em b), se fizermos $\ell_1 \to \ell_1 - 2\ell_2$, $\ell_3 \to \ell_3 - \ell_2$ e $\ell_4 \to \ell_4 + \ell_2$ e na matriz resultante expandirmos com respeito à primeira coluna, teremos:

$$\det \begin{bmatrix} 0 & 4 & -1 & -3 & 4 \\ 1 & 1 & 3 & 2 & -2 \\ 0 & 1 & -2 & -1 & 0 \\ 0 & 0 & 5 & -1 & 2 \\ 0 & 3 & -1 & 2 & 3 \end{bmatrix} = (-1)\det \begin{bmatrix} 4 & -1 & -3 & 4 \\ 1 & -2 & -1 & 0 \\ 0 & 5 & -1 & 2 \\ 3 & -1 & 2 & 3 \end{bmatrix}.$$

Nessa matriz faça $\ell_1 \to \ell_1 - 4\ell_2$ e $\ell_4 \to \ell_4 - 3\ell_2$ e obtemos

$$\det(B) = (-1)\det \begin{bmatrix} 0 & 7 & 1 & 4 \\ 1 & -2 & -1 & 0 \\ 0 & 5 & -1 & 2 \\ 0 & 5 & 5 & 3 \end{bmatrix} = (-1)(-1)\det \begin{bmatrix} 7 & 1 & 4 \\ 5 & -1 & 2 \\ 5 & 5 & 3 \end{bmatrix}$$

$$= \det \begin{bmatrix} 7 & 1 & 4 \\ 5 & -1 & 2 \\ 0 & 6 & 5 \end{bmatrix} = 24.$$

□

Transposta e Determinante

R6.4. Seja $A = [a_{ij}]$ uma matriz $n \times n$. Mostre que:

$$\det(A) = \sum_{i_1,\dots,i_n} \sigma(P_{i_1\dots i_n}) a_{i_1 1} a_{i_2 2} \cdots a_{i_n n}, \tag{6.1}$$

sendo $P_{i_1\dots i_n}$ uma matriz de permutação e $\sigma(P_{i_1\dots i_n}) = \pm 1$ o sinal desta matriz de permutação.

Solução: Seja $A = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \end{bmatrix} = [a_{ij}]$ uma matriz $n \times n$. E podemos escrever:

$$\begin{aligned}
\mathbf{c}_1 &= a_{11}\mathbf{e}_1 + a_{21}\mathbf{e}_2 + \cdots + a_{n1}\mathbf{e}_n \\
\mathbf{c}_2 &= a_{12}\mathbf{e}_1 + a_{22}\mathbf{e}_2 + \cdots + a_{n2}\mathbf{e}_n \\
\vdots \quad & \quad \vdots \\
\mathbf{c}_n &= a_{1n}\mathbf{e}_1 + a_{2n}\mathbf{e}_2 + \cdots + a_{nn}\mathbf{e}_n
\end{aligned}$$

sendo $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ a base canônica do $\mathbb{R}^n$. E, assim,

$$\begin{aligned}
\det(A) &= D(a_{11}\mathbf{e}_1 + \cdots + a_{n1}\mathbf{e}_n, \mathbf{c}_2, \dots, \mathbf{c}_n) \\
&= a_{11} D(\mathbf{e}_1, \mathbf{c}_2, \dots, \mathbf{c}_n) + \cdots + a_{1n} D(\mathbf{e}_n, \mathbf{c}_2, \dots, \mathbf{c}_n)
\end{aligned}$$

Se substituírmos $\mathbf{c}_2$ por $a_{12}\mathbf{e}_1 + \cdots + a_{n2}\mathbf{e}_n$ obteremos uma expressão semelhante, só que com mais termos. Feitas todas as substituições de $\mathbf{c}_2, \dots, \mathbf{c}_n$, e considerando que nos termos cujos índices têm repetições D é igual a zero, chegamos a expressão:

$$\begin{aligned}
\det(A) = D(\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n) &= \sum_{i_1,\dots,i_n} a_{i_1 1} a_{i_2 2} \cdots a_{i_n n} D(\mathbf{e}_{i_1}, \mathbf{e}_{i_2}, \dots, \mathbf{e}_{i_n}) \\
&= \sum_{i_1,\dots,i_n} a_{i_1 1} a_{i_2 2} \cdots a_{i_n n} \det(P_{i_1 i_2 \dots i_n}) \\
&= \sum_{i_1,\dots,i_n} \sigma(P_{i_1 i_2 \dots i_n}) a_{i_1 1} a_{i_2 2} \cdots a_{i_n n}.
\end{aligned}$$

No somatório acima i_l é diferente de todos os outros i_k se $l \neq k$. □

R6.5. Prove o teorema 6.14. Para todo $n > 1$, se A é uma matriz $n \times n$, então $\det(A) = \det(A^t)$.

Solução: Seja $A = [a_{ij}]$ e $B = [b_{ij}] = A^t$, portanto, $b_{ij} = a_{ji}$. Usando as equação para o determinante, deduzida no exercício R6.4, obtemos:

$$\begin{aligned}
\det(B) &= \sum_{i_1,\dots,i_n} \sigma(P_{i_1\dots i_n}) b_{i_1 1} b_{i_2 2} \cdots b_{i_n n} \\
&= \sum_{i_1,\dots,i_n} \det \begin{bmatrix} e_{i_1} & \cdots & e_{i_n} \end{bmatrix} a_{1 i_1} a_{2 i_2} \cdots a_{n i_n} \\
&= \sum_{j_1,\dots,j_n} \det \begin{bmatrix} e_{j_1} & \cdots & e_{j_n} \end{bmatrix}^t a_{j_1 1} a_{j_2 2} \cdots a_{j_n n} \\
&= \sum_{j_1,\dots,j_n} \det \begin{bmatrix} e_{j_1} & \cdots & e_{j_n} \end{bmatrix} a_{j_1 1} a_{j_2 2} \cdots a_{j_n n} \\
&= \sum_{j_1,\dots,j_n} \sigma(P_{j_1\dots j_n}) a_{j_1 1} a_{j_2 2} \cdots a_{j_n n} = \det(A).
\end{aligned}$$

Para entender melhor o que aconteceu na terceira igualdade veja a observação a seguir.

Dado um termo $a_{i_1 1} a_{i_2 2} \cdots a_{i_n n}$ do somatório do determinante que aparece no exercício R6.4. Observe que cada fator $a_{i_j k}$ tem dois índices i_j e k, por exemplo: $a_{i_2 2}$ tem o índice i_2 e o índice 2. Todos os valores do $\{1, 2, \dots, n\}$ aparecem no primeiro índice i_2. Então podemos reordenar os termos, de tal forma que o primeiro índice apareça com valores crescentes, desta forma podemos determinar $j_1, \dots, j_n$, tais que $a_{i_1 1} a_{i_2 2} \cdots a_{i_n n} = a_{1 j_1} a_{2 j_2} \cdots a_{n j_n}$. Por exemplo, na expressão do determinante de ordem 3×3 aparece o seguinte termo: $a_{31} a_{12} a_{23}$, que podemos reescrever da seguinte forma: $a_{12} a_{23} a_{31}$.

Além disso, a matriz $P_{j_1\dots j_n}$ é obtida quando tomamos $(P_{i_1\dots i_n})^t$, em que é possível verificar que $\sigma(P_{i_1\dots i_n}) = \sigma(P_{j_1\dots j_n})$. Por exemplo, associado ao termo $a_{31} a_{12} a_{23}$ temos a matriz de permutação P_{312}, como $a_{31} a_{12} a_{23} = a_{12} a_{23} a_{31}$, e obtemos a matriz P_{231} associado ao lado direito da igualdade. Agora verifique que $P_{231} = P_{312}^t$ e que $\det(P_{231}) = \det(P_{312}^t)$. $\square$

R6.6. Mostre que o determinante de qualquer matriz quadrada $A = [a_{ij}]$ pode ser calculando fazendo a expansão em qualquer de suas linhas ou colunas. Chamamos esta expansão de **Expansão de Laplace** do determinante. Assim, a expansão na i-ésima linha e na j-ésima coluna pode ser assim representada:

$$\begin{aligned}
\det(A) &= a_{i1}\Delta_{i1} + a_{i2}\Delta_{i2} + \cdots + a_{in}\Delta_{in} = \sum_{k=1}^{n} a_{ik}\Delta_{ik} \text{ e} \\
\det(A) &= a_{1j}\Delta_{1j} + a_{2j}\Delta_{2j} + \cdots + a_{nj}\Delta_{nj} = \sum_{l=1}^{n} a_{lj}\Delta_{lj}, \text{ respectivamente.}
\end{aligned}$$

Solução: Faremos a demonstração somente para a expansão pela j-ésima coluna. Considere a matriz escrita $A = \begin{bmatrix} \mathbf{c}_1 & \cdots & \mathbf{c}_{j-1} & \mathbf{c}_j & \mathbf{c}_{j+1} & \cdots & \mathbf{c}_n \end{bmatrix}$

como colunas, então:

$$
\begin{aligned}
\det(A) &= \det\begin{bmatrix}\mathbf{c}_1 & \cdots & \mathbf{c}_{j-1} & \mathbf{c}_j & \mathbf{c}_{j+1} & \cdots & \mathbf{c}_n\end{bmatrix} \\
&= D\left(\mathbf{c}_1, \ldots, \mathbf{c}_{j-1}, \mathbf{c}_j, \mathbf{c}_{j+1}, \ldots, \mathbf{c}_n\right) \\
&= -D\left(\mathbf{c}_1, \ldots, \mathbf{c}_j, \mathbf{c}_{j-1}, \mathbf{c}_{j+1}, \ldots, \mathbf{c}_n\right) \\
&\vdots \\
&= (-1)^{j-1} D\left(\mathbf{c}_j, \mathbf{c}_1, \ldots, \mathbf{c}_{j-1}, \mathbf{c}_{j+1}, \ldots, \mathbf{c}_n\right) \\
&= (-1)^{j-1}\det\begin{bmatrix}\mathbf{c}_j & \mathbf{c}_1 & \cdots & \mathbf{c}_{j-1} & \mathbf{c}_{j+1} & \cdots & \mathbf{c}_n\end{bmatrix}.
\end{aligned}
$$

Seja $B = \begin{bmatrix}\mathbf{c}_j & \mathbf{c}_1 & \cdots & \mathbf{c}_{j-1} & \mathbf{c}_{j+1} & \cdots & \mathbf{c}_n\end{bmatrix}$ e vamos calcular o determinante por fazer a expansão na primeira coluna de B

$$
\begin{aligned}
\det(A) &= (-1)^{j-1}\left(\sum_{k=1}^{n} a_{kj}(-1)^{k+1}\det(B_{k1})\right) \\
&= \sum_{k=1}^{n}(-1)^{k+1+j-1}a_{kj}\det(A_{kj}) \\
&= \sum_{k=1}^{n} a_{kj}(-1)^{k+j}\det(A_{kj}) = \sum_{k=1}^{n} a_{kj}\Delta_{kj}.
\end{aligned}
$$

No caso, da expansão do determinante com respeito a uma linha é só lembrar que o determinante também é linear antissimétrico com respeito às linhas da matriz. □

Determinantes e sistema de equações lineares

R6.7. Prove o teorema 6.17. O sistema (quadrado) $A\mathbf{x} = \mathbf{b}$ possui uma única solução se, e só se, $\det(A) \neq 0$. Nesse caso, a solução é dada por:

$$
x_1 = \frac{\det(A_1)}{\det(A)}, \quad x_2 = \frac{\det(A_2)}{\det(A)}, \quad \ldots, \quad x_n = \frac{\det(A_n)}{\det(A)}.
$$

Solução: Suponha que $\mathbf{x} = \sum_{j=1}^{n} x_j\mathbf{e}_j$ seja uma solução da equação e $\mathbf{e}_j$ seja os vetores da base canônica. Além disso, se $A = \begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n\end{bmatrix}$, logo $A\mathbf{x} = \sum_{j=1}^{n} x_j A\mathbf{e}_j = \sum_{j=1}^{n} x_j\mathbf{c}_j = \mathbf{b}$, isso se traduz por:

$$
\begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n\end{bmatrix}\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix} = \begin{bmatrix}b_1\\b_2\\\vdots\\b_n\end{bmatrix} \quad\Leftrightarrow\quad \textstyle\sum_{j=1}^{n} x_j\mathbf{c}_j = b \text{ e } b_i = \sum_{j=1}^{n} x_j a_{ij}.
$$

Assumindo que $\mathbf{x}$ é uma solução então:

$$
\begin{aligned}
A_k &= \begin{bmatrix}\mathbf{c}_1 & \cdots & \mathbf{c}_{k-1} & \mathbf{b} & \mathbf{c}_{k+1} & \cdots & \mathbf{c}_n\end{bmatrix} \\
&= \begin{bmatrix}\mathbf{c}_1 & \cdots & \mathbf{c}_{k-1} & \sum_{j=1}^{n} x_j\mathbf{c}_j & \mathbf{c}_{k+1} & \cdots & \mathbf{c}_n\end{bmatrix}
\end{aligned}
$$

e daí

$$\begin{aligned}
\det(A_k) &= D(\mathbf{c}_1, \ldots, \mathbf{c}_{k-1}, \sum_{j=1}^{n} x_j \mathbf{c}_j, \mathbf{c}_{k+1}, \ldots, \mathbf{c}_n) \\
&= \sum_{j=1}^{n} x_j D\left(\mathbf{c}_1, \ldots, \mathbf{c}_{k-1}, \mathbf{c}_j, \mathbf{c}_{k+1}, \ldots, \mathbf{c}_n\right) \\
&= x_k D\left(\mathbf{c}_1, \ldots, \mathbf{c}_{k-1}, \mathbf{c}_k, \mathbf{c}_{k+1}, \ldots, \mathbf{c}_n\right) \\
&= \det \begin{bmatrix} \mathbf{c}_1 & \cdots & \mathbf{c}_{k-1} & \mathbf{c}_k & \mathbf{c}_{k+1} & \cdots & \mathbf{c}_n \end{bmatrix} = x_k \det(A).
\end{aligned}$$

Logo, se $\det(A) \neq 0$ segue que:

$$x_k = \frac{\det(A_k)}{\det(A)}.$$

Como consequência se $\det(A) \neq 0$ a solução existe e é única. $\square$

Adjuntas clássicas e inversas

R6.8. Prove o teorema 6.19. Seja A uma matriz quadrada de ordem n qualquer. Então:

$$\operatorname{adj}(A)A = \det(A)I,$$

sendo I a matriz identidade. Assim, se $\det(A) \neq 0$,

$$A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A).$$

Solução: Se $\mathbf{x} = \begin{bmatrix} x_1 & \cdots & x_k & \cdots & x_n \end{bmatrix}^t \in \mathbb{R}^n$ é um vetor qualquer, e seja $A = [a_{ij}]$ vista como n colunas, isto é, $A = \begin{bmatrix} \mathbf{c}_1 & \cdots & \mathbf{c}_k & \cdots & \mathbf{c}_n \end{bmatrix}$, digamos que $\mathbf{u} = \begin{bmatrix} u_1 & \cdots & u_k & \cdots & u_n \end{bmatrix}^t = A\mathbf{x}$. Por fazer a expansão na k-ésima coluna (veja o exercício R6.6) de A_k, onde A_k é a matriz obtida por substituir $\mathbf{c}_k$ por $\mathbf{u}$, obtemos:

$$\det(A_k) = \sum_{i=1}^{n} (-1)^{i+k} u_i \det(A_{ik}), \text{ para todo } k = 1, 2, \ldots, n.$$

E pelo exercício R6.7 temos

$$x_k \det(A) = \sum_{i=1}^{n} (-1)^{i+k} \det(A_{ik}) u_i, \text{ para todo } k = 1, 2, \ldots, n.$$

Lembrando que $\Delta_{ik} = (-1)^{i+k} \det(A_{ik})$, podemos expressar estas n igualdades, usando a multiplicação de matriz por vetor, como a igualdade entre os vetores

$$\det(A) \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \Delta_{11} & \Delta_{21} & \cdots & \Delta_{n1} \\ \Delta_{12} & \Delta_{22} & \cdots & \Delta_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta_{1n} & \Delta_{2n} & \cdots & \Delta_{nn} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \operatorname{adj}(A)\mathbf{u}.$$

Como $\mathbf{u} = A\mathbf{x}$ temos que

$$\det(A)\mathbf{x} = \operatorname{adj}(A)\mathbf{u} = \operatorname{adj}(A)A\mathbf{x}, \quad \forall\, \mathbf{x} \in \mathbb{R}^n,$$

ou seja,

$$\operatorname{adj}(A)A = \det(A)I.$$

E, no caso de $\det(A) \neq 0$, temos que $\frac{1}{\det(A)} \operatorname{adj}(A) = A^{-1}$. □

R6.9. Prove o teorema 6.23. Seja B uma matriz quadrada triangular inferior (superior), em blocos, com os blocos diagonais $A_1, A_2, \ldots, A_k$, então,

$$\det(B) = \det(A_1)\det(A_2)\cdots\det(A_k).$$

Solução: Vamos demonstrar o resultado para um caso em particular. Considere a matriz:

$$B = \begin{bmatrix} a & b & a_{13} & a_{14} & a_{15} \\ c & d & a_{23} & a_{24} & a_{25} \\ 0 & 0 & e & f & g \\ 0 & 0 & h & i & j \\ 0 & 0 & k & l & m \end{bmatrix}.$$

Observe que A_1 é a matriz 2×2 e A_2 é a matriz 3×3. Vamos calcular o determinante de B por fazer, duas vezes, a expansão de Laplace na primeira linha, a seguir:

$$\begin{aligned} \det(B) &= a \det \begin{bmatrix} d & a_{23} & a_{24} & a_{25} \\ 0 & e & f & g \\ 0 & h & i & j \\ 0 & k & l & m \end{bmatrix} - c \det \begin{bmatrix} b & a_{13} & a_{14} & a_{15} \\ 0 & e & f & g \\ 0 & h & i & j \\ 0 & k & l & m \end{bmatrix} \\ &= ad \det \begin{bmatrix} e & f & g \\ h & i & j \\ k & l & m \end{bmatrix} - cb \det \begin{bmatrix} e & f & g \\ h & i & j \\ k & l & m \end{bmatrix} \\ &= (ad - bc) \det \begin{bmatrix} e & f & g \\ h & i & j \\ k & l & m \end{bmatrix} = \det(A_1)\det(A_2). \end{aligned}$$

O caso geral segue por raciocínio semelhante. □

R6.10. Seja $B = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & -1 \\ 0 & -2 & 1 \end{bmatrix}$. Encontre (a) $\operatorname{adj}(B)$, (b) $\det(B)$ e (c) B^{-1}, usando a $\operatorname{adj}(B)$.

Solução: Vamos começar por determinar $\operatorname{adj}(B)$. Para isto precisamos determinar todos os $\Delta_{ij} = (-1)^{i+j}\det(A_{ij})$, então:

$$\begin{aligned}\bar{B} &= \begin{bmatrix} (-1)^{1+1}\det(A_{11}) & (-1)^{1+2}\det(A_{12}) & (-1)^{1+3}\det(A_{13}) \\ (-1)^{2+1}\det(A_{21}) & (-1)^{2+2}\det(A_{22}) & (-1)^{2+3}\det(A_{23}) \\ (-1)^{3+1}\det(A_{31}) & (-1)^{3+2}\det(A_{32}) & (-1)^{3+3}\det(A_{33}) \end{bmatrix} \\ &= \begin{bmatrix} -1 & -1 & -2 \\ -1 & 1 & 2 \\ -1 & 1 & 0 \end{bmatrix}.\end{aligned}$$

E como sabemos que $\operatorname{adj}(B)B = \det(B)I$, calculando apenas a posição 11 da matriz $\operatorname{adj}(B)B$, obtemos o valor do determinante, então:

$$\operatorname{adj}(B)B = \bar{B}^t B = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 1 & 1 \\ -2 & 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & -1 \\ 0 & -2 & 1 \end{bmatrix} = \begin{bmatrix} -2 & & \\ & & \\ & & \end{bmatrix}.$$

Portanto, $\det(B) = -2$ e

$$B^{-1} = \begin{bmatrix} 1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & -1/2 \\ 1 & -1 & 0 \end{bmatrix}.$$

□

Exercícios propostos

P6.1. Use a regra de Cramer para calcular as soluções dos sistemas:

$$a)\begin{cases} 5x + 7y = 3 \\ 2x + 4y = 1 \end{cases} \qquad b)\begin{cases} 5x + 7y = 7 \\ -3x + z = -8 \\ y + 2z = -3 \end{cases} \qquad c)\begin{cases} 2x + y + z = 4 \\ -x + 2z = 2 \\ 3x + y + 3z = -2 \end{cases}$$

P6.2. Use a adjunta para calcular uma fórmula para a matriz inversa de $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

P6.3. Calcule $\det A$, sendo

$$\text{(a)} \quad A = \begin{bmatrix} 2 & 3 & 1 & -2 \\ 5 & 3 & 1 & 4 \\ 0 & 1 & 2 & -4 \\ 3 & -1 & -2 & 4 \end{bmatrix} \qquad \text{(b)} \quad A = \begin{bmatrix} 3 & -1 & 5 & 0 \\ 0 & 2 & 0 & 1 \\ 2 & 0 & -1 & 3 \\ 1 & 1 & 2 & 0 \end{bmatrix}$$

$$\text{(c)} \quad A = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ 19 & 18 & 0 & 0 & 0 \\ -6 & \pi & -5 & 0 & 0 \\ 4 & \sqrt{6} & 3 & 1/6 & 0 \\ 8 & 4 & \sqrt{7} & 5 & 4 \end{bmatrix}.$$

P6.4. Encontre A^{-1}, sendo

$$(a)\, A = \begin{bmatrix} 4 & -1 & 2 & -2 \\ 3 & -1 & 0 & 0 \\ 2 & 3 & 1 & 0 \\ 0 & 7 & 1 & 1 \end{bmatrix} \quad (b)\, A = \begin{bmatrix} 1 & 0 & x \\ 1 & 1 & x^2 \\ 2 & 2 & x^2 \end{bmatrix} \quad (c)\, A = \begin{bmatrix} -2 & -1 & 4 \\ 6 & -3 & -2 \\ 4 & 1 & 2 \end{bmatrix}$$

P6.5. Seja A uma matriz ortogonal, isto é, $A^tA = I$. Mostre que $\det(A) = \pm 1$.

P6.6. Mostre que

$$\det \begin{bmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{bmatrix} = (b-a)(c-a)(c-b).$$

P6.7. Seja V o espaço vetorial das matrizes $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ de ordem 2. Decida se a aplicação $D : V \to \mathbb{R}$ dada é bilinear (em relação as linhas)

$$(a) \quad D(M) = a + d, \quad (b) \quad D(M) = ac - bd$$
$$(c) \quad D(M) = ad, \quad (d) \quad D(M) = ab + cd.$$

P6.8. Sejam A, B, C e D matrizes quadradas de ordem n que comutam. Considere a matriz de ordem $2n$ em blocos $N = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$, então $\det(N) = \det(A)\det(D) - \det(B)\det(C)$. Mostre com um exemplo que a afirmação é falsa se as matrizes não comutarem.

P6.9. Determine uma fórmula para a área do triângulo cujos vértices são $\mathbf{0}$,$\mathbf{v}_1$ e $\mathbf{v}_2$ no $\mathbb{R}^2$.

P6.10. Seja R o triângulo com vértices $(x_1, y_1), (x_2, y_2)$ e (x_3, y_3). Mostre que

$$\{\text{área do triângulo}\} = \frac{1}{2} \left| \det \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{bmatrix} \right|.$$

P6.11. Encontre a área do paralelogramo cujos vértices são dados por:

$$a) \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 5 \\ 2 \end{smallmatrix}\right], \left[\begin{smallmatrix} 6 \\ 4 \end{smallmatrix}\right], \left[\begin{smallmatrix} 11 \\ 6 \end{smallmatrix}\right]; \quad b) \left[\begin{smallmatrix} -1 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ 5 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ -4 \end{smallmatrix}\right], \left[\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right]; \quad c) \left[\begin{smallmatrix} 0 \\ -2 \end{smallmatrix}\right], \left[\begin{smallmatrix} 6 \\ -1 \end{smallmatrix}\right], \left[\begin{smallmatrix} -3 \\ 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 3 \\ 2 \end{smallmatrix}\right].$$

P6.12. Determine o volume do paralelepípedo que tem um vértice na origem e os vértices adjacentes nos pontos $\begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$ e $\begin{bmatrix} 7 \\ 1 \\ 0 \end{bmatrix}$.

P6.13. Para cada uma das matrizes a seguir calcule (a) adj(A), (b) det(A) e (c) A^{-1}, usando a adj(A).

$$\text{(a)} \quad \begin{bmatrix} 1 & 3 & 2 \\ 2 & -4 & 3 \\ 5 & -2 & 1 \end{bmatrix} \qquad \text{(b)} \quad \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & -4 \\ 1 & 2 & 5 \end{bmatrix}.$$

Capítulo 7

Mudança de Base

Este capítulo é mais técnico, nele pretendemos explicar como as coordenadas de um vetor mudam, ao mudarmos de uma base para outra. Antes de começar é preciso fazer uma distinção sem a qual não é possível entender os conceitos discutidos neste capítulo. A distinção é a de que quando escrevemos $\mathbf{w}$ estamos imaginando um vetor (como ente geométrico) e quando escrevemos $[\mathbf{w}]$, estamos pensando nas coordenadas deste vetor com respeito à uma base.

7.1 Matriz Mudança de Coordenadas

Vamos considerar uma base $\alpha = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ do $\mathbb{R}^n$, então qualquer vetor $\mathbf{w} \in \mathbb{R}^n$ pode ser escrito de maneira única como combinação linear dos vetores de α, isto é, existem $x_1, x_2, \ldots, x_n \in \mathbb{R}$ tais que:

$$\mathbf{w} = x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + \cdots + x_n\mathbf{u}_n.$$

Se ordenarmos o conjunto α podemos associar para cada vetor $\mathbf{w}$ um único conjunto de números que informam as coordenadas do vetor $\mathbf{w}$, em relação aos vetores de α, e escrevemos $[\mathbf{w}]_\alpha = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^t$. Reciprocamente, se dermos um conjunto de n números $x_1, x_2, \ldots, x_n$ existe um único vetor associado, que é o vetor $\mathbf{w} = x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + \cdots + x_n\mathbf{u}_n$.

Agora se tivermos outra base, digamos $\beta = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, então podemos encontrar $y_1, y_2, \ldots, y_n \in \mathbb{R}$, tais que o mesmo vetor $\mathbf{w}$ se escreve

$$\mathbf{w} = y_1\mathbf{v}_1 + y_2\mathbf{v}_2 + \cdots + y_n\mathbf{v}_n,$$

e as coordenadas desse vetor com respeito à base β são $[\mathbf{w}]_\beta = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^t$. Nesta seção vamos entender como relacionar as coordenadas de $\mathbf{w}$, na base α, com as coordenadas de $\mathbf{w}$, com respeito à base β.

7.1.1 Dimensão 2

Faremos as contas somente para o caso em que a dimensão do espaço vetorial é 2, isso porque o resultado obtido em dimensão 2 estende-se para espaços de dimensão maior sem nenhuma dificuldade.

Sejam $\alpha = \{\mathbf{u}_1, \mathbf{u}_2\}$ e $\beta = \{\mathbf{v}_1, \mathbf{v}_2\}$ duas bases ordenadas de $\mathbb{R}^2$. Então dado, o vetor

$$\begin{aligned}\mathbf{w} &= x_1\mathbf{u}_1 + x_2\mathbf{u}_2\\ &= y_1\mathbf{v}_1 + y_2\mathbf{v}_2.\end{aligned}$$

Como $\mathbf{v}_1$ e $\mathbf{v}_2$ são vetores, podemos determinar as coordenadas destes vetores com respeito à base α, isto é,

$$\begin{aligned}\mathbf{v}_1 &= a_{11}\mathbf{u}_1 + a_{21}\mathbf{u}_2\\ \mathbf{v}_2 &= a_{12}\mathbf{u}_1 + a_{22}\mathbf{u}_2.\end{aligned}$$

Substituindo na igualdade acima obtemos:

$$\begin{aligned}\mathbf{w} &= y_1\mathbf{v}_1 + y_2\mathbf{v}_2\\ &= y_1(a_{11}\mathbf{u}_1 + a_{21}\mathbf{u}_2) + y_2(a_{12}\mathbf{u}_1 + a_{22}\mathbf{u}_2)\\ &= (a_{11}y_1 + a_{12}y_2)\mathbf{u}_1 + (a_{21}y_1 + a_{22}y_2)\mathbf{u}_2.\end{aligned}$$

E usando a unicidade da representação de um vetor em termos de uma base ordenada e a multiplicação de matrizes obtemos que:

$$\begin{bmatrix}x_1\\x_2\end{bmatrix} = \begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix}\begin{bmatrix}y_1\\y_2\end{bmatrix}.$$

Observe que qualquer que seja o vetor $\mathbf{w}$, se soubermos as coordenadas dele com respeito à base β, $[\mathbf{w}]_\beta$, podemos encontrar as coordenadas de $\mathbf{w}$ na base α, $[\mathbf{w}]_\alpha$, bastando para isso multiplicar $[\mathbf{w}]_\beta$ pela matriz acima. Chamamos essa matriz de **matriz de mudança de coordenadas** da base β para a base α e a denotamos por $[I]^\beta_\alpha$.

Exemplo 7.1
Considere $V = \mathbb{R}^2$ e $\alpha = \left\{\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]\right\}$ e a base $\beta = \left\{\left[\begin{smallmatrix}-1\\1\end{smallmatrix}\right], \left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]\right\}$. Então, para determinarmos a matriz mudança da base α para a base β, $[I]^\beta_\alpha$, precisamos encontrar as coordenadas dos vetores da base β com respeito à base α, isto é,

$$\begin{bmatrix}-1\\1\end{bmatrix} = a_{11}\begin{bmatrix}1\\0\end{bmatrix} + a_{21}\begin{bmatrix}0\\1\end{bmatrix} \text{ e também } \begin{bmatrix}1\\1\end{bmatrix} = a_{12}\begin{bmatrix}1\\0\end{bmatrix} + a_{22}\begin{bmatrix}0\\1\end{bmatrix}.$$

Essas equações vetoriais são muito fáceis de serem resolvidas e as soluções são $a_{11} = -1$, $a_{12} = 1$, $a_{21} = 1$ e $a_{22} = 1$. Portanto,

$$[I]^\beta_\alpha = \begin{bmatrix}-1 & 1\\ 1 & 1\end{bmatrix}.$$

Vamos determinar também $[I]^\alpha_\beta$. Para isso precisamos escrever os vetores da base α, em termos da base β. Depois de fazermos as contas chegamos que:

$$\begin{bmatrix}1\\0\end{bmatrix} = -\,1/2\begin{bmatrix}-1\\1\end{bmatrix} + 1/2\begin{bmatrix}1\\1\end{bmatrix} \text{ e } \begin{bmatrix}0\\1\end{bmatrix} = 1/2\begin{bmatrix}-1\\1\end{bmatrix} + 1/2\begin{bmatrix}1\\1\end{bmatrix}.$$

Portanto,

$$[I]_\beta^\alpha = \begin{bmatrix} -1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}.$$

Vamos fazer duas observação: a primeira é a de que $[I]_\beta^\alpha [I]_\alpha^\beta = I$, portanto, uma é inversa da outra(este resultado é geral!) A segunda é a de que a matriz $[I]_\alpha^\beta$ tem como colunas os vetores da base β, esse fato sempre ocorre se a base de chegada é a base canônica.

7.1.2 Caso Geral

Vamos tratar o caso geral. Suponha que $\alpha = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ é uma base de V assim como $\beta = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ e, que $\mathbf{w}$ seja um vetor qualquer de V, então podemos determinar as coordenadas

$$[\mathbf{w}]_\alpha = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \text{ e } [\mathbf{w}]_\beta = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

Escrevendo os vetores da base β, em termos da base α, podemos determinar os n^2 números a_{ij} a seguir:

$$\begin{aligned}
\mathbf{v}_1 &= a_{11}\mathbf{u}_1 + a_{21}\mathbf{u}_2 + \cdots + a_{n1}\mathbf{u}_n \\
\mathbf{v}_2 &= a_{12}\mathbf{u}_1 + a_{22}\mathbf{u}_2 + \cdots + a_{n2}\mathbf{u}_n \\
&\cdots\cdots \\
\mathbf{v}_j &= a_{1j}\mathbf{u}_1 + a_{2j}\mathbf{u}_2 + \cdots + a_{nj}\mathbf{u}_n \\
&\cdots\cdots \\
\mathbf{v}_n &= a_{1n}\mathbf{u}_1 + a_{2n}\mathbf{u}_2 + \cdots + a_{nn}\mathbf{u}_n
\end{aligned}$$

E, montando a matriz,

$$[I]_\alpha^\beta = \begin{bmatrix} [\mathbf{v}_1]_\alpha & [\mathbf{v}_2]_\alpha & \cdots & [\mathbf{v}_n]_\alpha \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix}.$$

Essa matriz foi obtida ao pegar os coeficientes que aparecem na expressão do vetor $\mathbf{v}_j$ como combinação linear dos vetores $\mathbf{u}_{i's}$ e colocar na j-ésima coluna. Essa matriz é chamada de **matriz de mudança da coordenadas de β para a base α** ou simplesmente **matriz de mudança de coordenadas**.

Observação 7.2
Na literatura a matriz $[I]_\alpha^\beta$ é muitas vezes chamada de matriz de mudança da base α para a base β, ou, simplesmente, matriz de mudança de base. Você poderia pensar que cometemos um equivoco, mas, de fato, não cometemos. Mais para frente justificaremos este nome.

7.2 Aplicações lineares e Matrizes

Até o momento sabemos que dada uma aplicação linear $T : \mathbb{R}^n \to \mathbb{R}^m$ podemos associar uma matriz $A = [a_{ij}]$, de ordem $m \times n$ e, reciprocamente, dada uma matriz $A = [a_{ij}]$, de ordem $m \times n$, podemos determinar uma aplicação linear $S : \mathbb{R}^n \to \mathbb{R}^m$ definida por $\mathbf{x} \in \mathbb{R}^n \mapsto A\mathbf{x} \in \mathbb{R}^m$. Vamos estender esse conceito para quando levarmos em conta as coordenadas de um vetor.

Vamos iniciar considerando a aplicação linear $T : \mathbb{R}^n \to \mathbb{R}^m$, e sejam $\alpha = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ uma base de $\mathbb{R}^n$ e $\beta = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ uma base de $\mathbb{R}^m$. Observe que em qualquer vetor $\mathbf{u} \in \mathbb{R}^n$ o vetor $T(u) \in \mathbb{R}^m$ podemos determinar as coordenadas do mesmo com respeito à base β em particular,

$$\begin{aligned}
T(\mathbf{u}_1) &= a_{11}\mathbf{v}_1 + a_{21}\mathbf{v}_2 + \cdots + a_{m1}\mathbf{v}_m \\
T(\mathbf{u}_2) &= a_{12}\mathbf{v}_1 + a_{22}\mathbf{v}_2 + \cdots + a_{m2}\mathbf{v}_m \\
&\cdots\cdots\cdots \\
T(\mathbf{u}_j) &= a_{1j}\mathbf{v}_1 + a_{2j}\mathbf{v}_2 + \cdots + a_{mj}\mathbf{v}_m \\
&\cdots\cdots\cdots \\
T(\mathbf{u}_n) &= a_{1n}\mathbf{v}_1 + a_{2n}\mathbf{v}_2 + \cdots + a_{mn}\mathbf{v}_m
\end{aligned}$$

Assim, podemos associar a matriz

$$[T]^{\alpha}_{\beta} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

que é chamada **matriz da aplicação linear T com respeito às bases α e β.**

Exemplo 7.3
Considere $T : \mathbb{R}^3 \to \mathbb{R}^2$ definida por

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} 2x + y - z \\ 3x - 3y - 4z \end{bmatrix} \text{ e as bases}$$

$$\alpha = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\} \text{ e } \beta = \left\{ \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}.$$

Encontre a matriz de T com respeito à estas bases. Precisamos encontrar as coordenadas, na base β, dos vetores da base α avaliados por T.

$$\begin{aligned} T\left(\begin{bmatrix}1\\1\\1\end{bmatrix}\right) &= \begin{bmatrix}2\\-4\end{bmatrix} = (-3)\begin{bmatrix}-1\\1\end{bmatrix} + (-1)\begin{bmatrix}1\\1\end{bmatrix};\\ T\left(\begin{bmatrix}-1\\-1\\0\end{bmatrix}\right) &= \begin{bmatrix}-3\\0\end{bmatrix} = (3/2)\begin{bmatrix}-1\\1\end{bmatrix} + (-3/2)\begin{bmatrix}1\\1\end{bmatrix};\\ T\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\right) &= \begin{bmatrix}2\\3\end{bmatrix} = (1/2)\begin{bmatrix}-1\\1\end{bmatrix} + (5/2)\begin{bmatrix}1\\1\end{bmatrix}. \end{aligned}$$

Portanto, a matriz de T, com respeito às bases α e β é

$$[T]_\beta^\alpha = \begin{bmatrix}-3 & 3/2 & 1/2\\ -1 & -3/2 & 5/2\end{bmatrix}.$$

Teorema 7.4
Sejam $T : \mathbb{R}^n \to \mathbb{R}^m$ uma aplicação linear, α uma base de $\mathbb{R}^n$, β uma base de $\mathbb{R}^m$, então

$$[T(\mathbf{w})]_\beta = [T]_\beta^\alpha[\mathbf{w}]_\alpha.$$

Demonstração: Esse teorema nos diz que se tomarmos o vetor $\mathbf{w}$ e calcularmos as coordenadas de $T(\mathbf{w})$, com respeito à base β, será o mesmo que calcularmos $[\mathbf{w}]_\alpha$ vezes a matriz da aplicação T, com respeito às bases α e β.

Faremos a demonstração somente para o caso $n = 2$ e $m = 3$, por acreditar que isso é bem mais instrutivo que a demonstração no caso geral. Para começar, sejam $\alpha = \{\mathbf{u}_1, \mathbf{u}_2\}$ uma base do $\mathbb{R}^2$ e $\beta = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ uma base do $\mathbb{R}^3$. Sabemos que existem únicos coeficientes $a_{ij} \in \mathbb{R}$, tais que:

$$\begin{aligned} T(\mathbf{u}_1) &= a_{11}\mathbf{v}_1 + a_{21}\mathbf{v}_2 + a_{31}\mathbf{v}_3\\ T(\mathbf{u}_2) &= a_{12}\mathbf{v}_1 + a_{22}\mathbf{v}_2 + a_{32}\mathbf{v}_3, \end{aligned}$$

e obtemos a matriz

$$[T]_\beta^\alpha = \begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\\ a_{31} & a_{32}\end{bmatrix}.$$

Seja $\mathbf{w}$ um vetor de $\mathbb{R}^2$ e sejam $[\mathbf{w}]_\alpha = \left[\begin{smallmatrix}x_1\\x_2\end{smallmatrix}\right]$ e $[T(\mathbf{w})]_\beta = \left[\begin{smallmatrix}y_1\\y_2\\y_3\end{smallmatrix}\right]$ as suas coordenadas.

Logo, $\mathbf{w} = x_1\mathbf{u}_1 + x_2\mathbf{u}_2$ e por fazer uso da linearidade da aplicação T temos:

$$\begin{aligned} T(\mathbf{w}) &= T(x_1\mathbf{u}_1 + x_2\mathbf{u}_2)\\ &= x_1T(\mathbf{u}_1) + x_2T(\mathbf{u}_2)\\ &= x_1(a_{11}\mathbf{v}_1 + a_{21}\mathbf{v}_2 + a_{31}\mathbf{v}_3) + x_2(a_{12}\mathbf{v}_1 + a_{22}\mathbf{v}_2 + a_{32}\mathbf{v}_3)\\ &= (a_{11}x_1 + a_{12}x_2)\mathbf{v}_1 + (a_{21}x_1 + a_{22}x_2)\mathbf{v}_2 + (a_{31}x_1 + a_{32}x_2)\mathbf{v}_3. \end{aligned}$$

Mas $T(\mathbf{w}) = y_1\mathbf{v}_1 + y_2\mathbf{v}_2 + y_3\mathbf{v}_3$ e como as coordenadas com respeito a uma base são únicas segue que

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 \\ y_2 = a_{21}x_1 + a_{22}x_2 \\ y_3 = a_{31}x_1 + a_{32}x_2 \end{cases}, \text{ que é equivalente a, } \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Isto é, $[T(\mathbf{w})]_\beta = [T]^\alpha_\beta [\mathbf{w}]_\alpha$. □

Exemplo 7.5
Considere o caso especial $I : \mathbb{R}^n \to \mathbb{R}^n$ o operador identidade, isto é, $I(\mathbf{v}) = \mathbf{v}$ para todo $\mathbf{v} \in \mathbb{R}^n$. Considere $\alpha = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ uma base do domínio de I e uma base $\beta = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ do contradomínio de I. Vamos determinar a matriz deste operador com respeito à estas bases. Para isso calcule:

$$\begin{aligned} I(\mathbf{v}_1) &= \mathbf{v}_1 = a_{11}\mathbf{u}_1 + a_{21}\mathbf{u}_2 + \cdots + a_{n1}\mathbf{u}_n; \\ I(\mathbf{v}_2) &= \mathbf{v}_2 = a_{12}\mathbf{u}_1 + a_{22}\mathbf{u}_2 + \cdots + a_{n2}\mathbf{u}_n; \\ &\cdots\cdots\cdots \\ I(\mathbf{v}_j) &= \mathbf{v}_j = a_{1j}\mathbf{u}_1 + a_{2j}\mathbf{u}_2 + \cdots + a_{nj}\mathbf{u}_n; \\ &\cdots\cdots\cdots \\ I(\mathbf{v}_n) &= \mathbf{v}_n = a_{1n}\mathbf{u}_1 + a_{2n}\mathbf{u}_2 + \cdots + a_{nn}\mathbf{u}_n. \end{aligned}$$

Observe que a matriz de I com respeito às duas bases é obtida por pegar as coordenadas do vetor $\mathbf{v}_j$, em termos da base β e colocar na j-ésima coluna da matriz. Mas isso é exatamente a forma de calcular a matriz de mudança de coordenada. Portanto,

$$[I]^\alpha_\beta = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}.$$

E por isso denotar a matriz de mudança de coordenadas da base α para a base β por $[I]^\alpha_\beta$ não é nada de especial, é apenas a matriz do operador I com respeito às bases α e β.

Teorema 7.6
Sejam $T : \mathbb{R}^n \to \mathbb{R}^m$ e $S : \mathbb{R}^m \to \mathbb{R}^k$ duas aplicações lineares e α, β e γ bases de $\mathbb{R}^n, \mathbb{R}^m$ e $\mathbb{R}^k$, respectivamente. Então, a composta de $S \circ T : \mathbb{R}^n \to \mathbb{R}^k$, é linear e

$$[S \circ T]^\alpha_\gamma = [S]^\beta_\gamma \cdot [T]^\alpha_\beta.$$

A demonstração desse resultado é fácil, mas muito trabalhosa, veja o exercício R7.2.

Observação 7.7
Você há de convir que seria muito mais natural definir a multiplicação entre matrizes como a simples multiplicação entre as entradas correspondentes e somente

para matrizes de mesmo tamanho, similarmente ao que ocorre com a operação de soma de matrizes. Definimos dessa forma para tornar o Teorema 7.6 verdadeiro.

O Teorema 7.6 nos diz que a multiplicação entre matrizes é compatível com a composição de funções lineares.

Corolário 7.8
Se $T : \mathbb{R}^n \to \mathbb{R}^n$ é um operador linear invertível e se α e β são bases do domínio e do contradomínio, respectivamente. Logo, $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ também é um operador linear e

$$[T^{-1}]_\alpha^\beta = ([T]_\beta^\alpha)^{-1}.$$

Demonstração: Segue das seguintes duas observações. Em primeiro lugar, se $I : \mathbb{R}^n \to \mathbb{R}^n$ é o operador identidade, e γ é uma base de qualquer de $\mathbb{R}^n$, então, a0 calcularmos $[I]_\gamma^\gamma$ obtemos sempre a matriz identidade, qualquer que seja γ escolhida. A segunda observação é a seguinte

$$[T^{-1}]_\alpha^\beta [T]_\beta^\alpha = [T^{-1} \circ T]_\alpha^\alpha = [I]_\alpha^\alpha.$$

Portanto, $[T^{-1}]_\alpha^\beta = ([T]_\beta^\alpha)^{-1}$. □

Segue deste corolário que se $[I]_\alpha^\beta$ é a matriz de mudança de coordenadas da base β para a base α, então $[I]_\beta^\alpha = ([I]_\alpha^\beta)^{-1}$.

Se $T : \mathbb{R}^n \to \mathbb{R}^m$ é uma aplicação linear e, α e α' são bases de $\mathbb{R}^n$ e β e β' são bases de $\mathbb{R}^m$, então podemos associar as seguintes matrizes $[T]_\beta^\alpha$ e $[T]_{\beta'}^{\alpha'}$ à aplicação linear T.

Como podemos relacionar estas matrizes? Como elas provêm da mesma transformação linear, devem ter a mesma ação sobre os vetores de $\mathbb{R}^n$, o que deve sofrer alteração, são as coordenadas desses vetores.

Observe que para avaliar o vetor $\mathbf{w} \in \mathbb{R}^n$ em T, por usar a matriz $[T]_\beta^\alpha$, precisamos obter as coordenadas de $\mathbf{w}$ na base α, digamos ainda que conheçamos as coordenadas na base α', isto é, $[\mathbf{w}]_{\alpha'}$. Portanto, para obtermos as coordenadas na base α precisamos da matriz $[I]_\alpha^{\alpha'}$ e, por um lado,

$$[T(\mathbf{w})]_{\beta'} = [T]_{\beta'}^{\alpha'} [\mathbf{w}]_{\alpha'}$$

e por outro,

$$[T(\mathbf{w})]_\beta = [T]_\beta^\alpha [I]_\alpha^{\alpha'} [\mathbf{w}]_{\alpha'}.$$

Se ainda conhecermos a matriz $[I]_\beta^{\beta'}$, então temos a igualdade

$$[I]_\beta^{\beta'} [T]_{\beta'}^{\alpha'} [\mathbf{w}]_{\alpha'} = [I]_\beta^{\beta'} [T(\mathbf{w})]_{\beta'} = [T(\mathbf{w})]_\beta = [T]_\beta^\alpha [I]_\alpha^{\alpha'} [\mathbf{w}]_{\alpha'}.$$

Como isso vale para todo vetor $\mathbf{w}$, logo essa igualdade é válida entre as matrizes, isto é,

$$[I]_\beta^{\beta'} [T]_{\beta'}^{\alpha'} = [T]_\beta^\alpha [I]_\alpha^{\alpha'}.$$

Exemplo 7.9
Considere a mesma aplicação $T : \mathbb{R}^3 \to \mathbb{R}^2$ definida no exemplo 7.3 e α' a base canônica do $\mathbb{R}^3$ e β' a base canônica do $\mathbb{R}^2$, então

$$[T]_{\beta'}^{\alpha'} = \begin{bmatrix} 2 & 1 & -1 \\ 3 & -3 & -4 \end{bmatrix}.$$

Vamos conectar com a matriz

$$[T]_{\beta}^{\alpha} = \begin{bmatrix} -3 & 3/2 & 1/2 \\ -1 & -3/2 & 5/2 \end{bmatrix},$$

calculada no exemplo 7.3 com essa matriz, para isso precisamos das matrizes $[I]_{\beta}^{\beta'}$, que foi calculada no exemplo 7.1, e também $[I]_{\alpha}^{\alpha'}$. Executando as contas obtemos:

$$[I]_{\beta}^{\beta'} = \begin{bmatrix} -1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} \text{ e } [I]_{\alpha}^{\alpha'} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 1 \\ 1 & -1 & 0 \end{bmatrix}.$$

Vale a seguinte igualdade (verifique):

$$\begin{aligned} \begin{bmatrix} -1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 2 & 1 & -1 \\ 3 & -3 & -4 \end{bmatrix} &= [I]_{\beta}^{\beta'}[T]_{\beta'}^{\alpha'} \\ &= \begin{bmatrix} 1/2 & -2 & -3/2 \\ 5/2 & -1 & -5/2 \end{bmatrix} = [T]_{\beta}^{\alpha}[I]_{\alpha}^{\alpha'} \\ &= \begin{bmatrix} -3 & 3/2 & 1/2 \\ -1 & -3/2 & 5/2 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 1 \\ 1 & -1 & 0 \end{bmatrix}. \end{aligned}$$

Se na equação $[I]_{\beta}^{\beta'}[T]_{\beta'}^{\alpha'} = [T]_{\beta}^{\alpha}[I]_{\alpha}^{\alpha'}$ tivermos $m = n$ e $\alpha = \beta$ e $\alpha' = \beta'$, a igualdade aqui demonstrada se torna $[I]_{\alpha}^{\alpha'}[T]_{\alpha'}^{\alpha'} = [T]_{\alpha}^{\alpha}[I]_{\alpha}^{\alpha'}$ e, lembrando que $([I]_{\alpha}^{\alpha'})^{-1} = [I]_{\alpha'}^{\alpha}$ se multiplicarmos a igualdade por $([I]_{\alpha}^{\alpha'})^{-1}$ obtemos:

$$[T]_{\alpha'}^{\alpha'} = [I]_{\alpha'}^{\alpha}[T]_{\alpha}^{\alpha}[I]_{\alpha}^{\alpha'}.$$

Observação 7.10
Continuando com a notação anterior, observe que ao fazemos $[T]_{\alpha}^{\alpha}[I]_{\alpha}^{\alpha'}$ obteremos $[T]_{\alpha}^{\alpha'}$. Portanto, a matriz $[I]_{\alpha}^{\alpha'}$ toma a matriz de T, na base α, α e retorna a matriz de T com respeito às bases α', α, podemos dizer que a matriz $[I]_{\alpha}^{\alpha'}$ é a matriz de mudança da base α para a base α'. Isso justifica o nome dado anteriormente na observação 7.2.

Definição 7.11
Sejam A e B duas matrizes quadradas. Dizemos que A e B são matrizes semelhantes e denotamos por $A \cong B$ se existe uma matriz P invertível, tal que

$$B = P^{-1}AP.$$

Disso temos que todas as matrizes associadas a um operador são semelhantes.

Exercícios resolvidos

R7.1. Sejam $\alpha = \{\mathbf{u}_1, \mathbf{u}_2\}$ e $\beta = \{\mathbf{v}_1, \mathbf{v}_2\}$ as bases de um espaço vetorial V, e suponha que $\mathbf{v}_1 = 6\mathbf{u}_1 - 2\mathbf{u}_2$ e $\mathbf{v}_2 = 9\mathbf{u}_1 - 4\mathbf{u}_2$.

a) Determine a matriz de mudança de coordenadas $[I]^\alpha_\beta$;

b) Determine, usando o item a), $[\mathbf{w}]_\beta$ para $\mathbf{w} = 3\mathbf{u}_1 - 4\mathbf{u}_2$.

Solução: a) Da expressão de $\mathbf{v}_1$ e $\mathbf{v}_2$ como combinação linear de $\mathbf{u}_1, \mathbf{u}_2$ obtemos $[I]^\beta_\alpha = \begin{bmatrix} 6 & 9 \\ -2 & -4 \end{bmatrix}$. Como $\left([I]^\beta_\alpha\right)^{-1} = [I]^\alpha_\beta$ segue que

$$[I]^\alpha_\beta = \begin{bmatrix} 2/3 & 3/2 \\ -1/3 & -1 \end{bmatrix}.$$

b) Como sabemos que as coordenadas de $\mathbf{w}$, com respeito à base α, para encontrarmos as coordenadas de $\mathbf{w}$, com respeito à base β, basta calcularmos:

$$[\mathbf{w}]_\beta = [I]^\alpha_\beta [\mathbf{w}]_\alpha = \begin{bmatrix} 2/3 & 3/2 \\ -1/3 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ -4 \end{bmatrix} = \begin{bmatrix} -4 \\ 3 \end{bmatrix}.$$

□

R7.2. Prove o teorema 7.6. Sejam $T : \mathbb{R}^n \to \mathbb{R}^m$ e $S : \mathbb{R}^m \to \mathbb{R}^k$ duas aplicações lineares e α, β e γ bases de $\mathbb{R}^n, \mathbb{R}^m$ e $\mathbb{R}^k$, respectivamente. Então, a composta de $S \circ T : \mathbb{R}^n \to \mathbb{R}^k$ é linear e

$$[S \circ T]^\alpha_\gamma = [S]^\beta_\gamma \cdot [T]^\alpha_\gamma.$$

Solução: Suponha que $\alpha = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\} \subset \mathbb{R}^n$, $\beta = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\} \subset \mathbb{R}^m$ e $\gamma = \{\mathbf{w}_1, \ldots, \mathbf{w}_k\} \subset \mathbb{R}^k$ sejam bases de $\mathbb{R}^n$, $\mathbb{R}^m$ e $\mathbb{R}^k$, respectivamente. Podemos determinar escalares a_{ij} e b_{jl}, satisfazendo:

$$T(\mathbf{u}_i) = \sum_{j=1}^m a_{ij}\mathbf{v}_j \text{ e } S(\mathbf{v}_j) = \sum_{l=1}^k b_{jl}\mathbf{w}_l, \text{ com } i = 1, \ldots, n \text{ e } j = 1, \ldots, m.$$

E isto determina as matrizes $[S]^\beta_\gamma = [b_{jl}]_{k\times m}$ e $[T]^\alpha_\beta = [a_{ij}]_{m\times n}$. Agora, observe o seguinte:

$$\begin{aligned}(S \circ T)(\mathbf{u}_i) = S(T(\mathbf{u}_i)) &= S\left(\sum_{j=1}^m a_{ij}\mathbf{v}_j\right) \\ &= \sum_{j=1}^m a_{ij} S(\mathbf{v}_j) \\ &= \sum_{j=1}^m a_{ij} \sum_{l=1}^k b_{jl}\mathbf{w}_l = \sum_{l=1}^k \left(\sum_{j=1}^m a_{ij} b_{jl}\right)\mathbf{w}_l.\end{aligned}$$

O escalar $\sum_{j=1}^{m} a_{ij}b_{jl}$ é a entrada il da matriz da transformação linear $S \circ T$ com respeito às bases α e γ. Por outro lado, a entrada na posição il da matriz obtida por multiplicar $[b_{jl}]_{k\times m}$ por $[a_{ij}]_{m\times n}$ é o escalar $\sum_{j=1}^{m} a_{ij}b_{jl}$. De onde obtemos a igualdade desejada. $\square$

R7.3. Considere o operador linear F de $\mathbb{R}^2$ definido por $F\left(\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]\right) = \left[\begin{smallmatrix} 5x-y \\ 2x+y \end{smallmatrix}\right]$ e as bases de $\mathbb{R}^2$ a seguir:

$$\alpha = \left\{\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]\right\} \text{ e } \beta = \left\{\left[\begin{smallmatrix} 1 \\ 4 \end{smallmatrix}\right], \left[\begin{smallmatrix} 2 \\ 7 \end{smallmatrix}\right]\right\}.$$

a) Encontre a matriz P de mudança de coordenada da base α para a base β e a matriz Q de mudança de coordenada da base β para a base α.

b) Encontre a matriz A que representa F na base α.

c) Encontre a matriz B que representa F na base β.

Solução: a) Vamos começar determinando a matriz $Q = [I]_{\alpha}^{\beta}$, a qual é muito fácil de determinar, visto que precisamos escrever os vetores da base β como combinação linear dos vetores da base α que é a base canônica, além disso, $P = Q^{-1}$, então:

$$Q = \begin{bmatrix} 1 & 2 \\ 4 & 7 \end{bmatrix} \text{ logo } P = Q^{-1} = Q = \frac{1}{-1}\begin{bmatrix} 7 & -2 \\ -4 & 1 \end{bmatrix} = \begin{bmatrix} -7 & 2 \\ 4 & -1 \end{bmatrix}.$$

b) A matriz $A = [T]_{\alpha}^{\alpha}$ é facilmente obtida da expressão, basta fazer $\left[\begin{smallmatrix} 5x-y \\ 2x+y \end{smallmatrix}\right] = x\left[\begin{smallmatrix} 5 \\ 2 \end{smallmatrix}\right] + y\left[\begin{smallmatrix} -1 \\ 1 \end{smallmatrix}\right]$ e daí

$$A = [T]_{\alpha}^{\alpha} = \begin{bmatrix} 5 & -1 \\ 2 & 1 \end{bmatrix}.$$

c) Como sabemos que $[T]_{\beta}^{\beta} = [I]_{\beta}^{\alpha}[T]_{\alpha}^{\alpha}[I]_{\alpha}^{\beta}$, logo:

$$[T]_{\beta}^{\beta} = \begin{bmatrix} -7 & 2 \\ 4 & -1 \end{bmatrix}\begin{bmatrix} 5 & -1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 4 & 7 \end{bmatrix} = \begin{bmatrix} -7 & 2 \\ 4 & -1 \end{bmatrix}\begin{bmatrix} 1 & 3 \\ 6 & 11 \end{bmatrix} = \begin{bmatrix} 5 & 1 \\ -2 & 1 \end{bmatrix}.$$

$\square$

R7.4. Considere $T : \mathbb{R}^3 \to \mathbb{R}^3$ uma transformação linear, tal que $T\left(\left[\begin{smallmatrix} 1 \\ 0 \\ 1 \end{smallmatrix}\right]\right) = \left[\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix}\right]$, $T\left(\left[\begin{smallmatrix} 0 \\ 1 \\ -2 \end{smallmatrix}\right]\right) = \left[\begin{smallmatrix} 0 \\ 1 \\ 0 \end{smallmatrix}\right]$ e $T\left(\left[\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right]\right) = \left[\begin{smallmatrix} 0 \\ 0 \\ 1 \end{smallmatrix}\right]$.

a) Mostre que $\beta = \left\{\left[\begin{smallmatrix} 1 \\ 0 \\ 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ 1 \\ -2 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right]\right\}$ é uma base do $\mathbb{R}^3$.

b) Determine $[v]_{\beta}$ se $v = \left[\begin{smallmatrix} 2 \\ -1 \\ 0 \end{smallmatrix}\right]$.

c) Determine $T\left(\left[\begin{smallmatrix} 2 \\ -1 \\ 0 \end{smallmatrix}\right]\right)$.

Solução: a) Vamos montar uma matriz A, por colocar os vetores da base β nas colunas e então

$$\det(A) = \det \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & -2 & 1 \end{bmatrix} = \det \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & -2 & 0 \end{bmatrix} = \det \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix} = 2 \neq 0,$$

E isso garante que estes vetores são linearmente independentes, uma vez que, se os vetores não fossem LI, então eles estariam em um plano, nesse caso o volume do paralelepípedo determinado por esses vetores seria 0, o que não ocorre.

b) Se α é a base canônica do $\mathbb{R}^3$ precisamos encontrar a matriz $[I]^{\alpha}_{\beta}$, mas calcular a matriz $[I]^{\beta}_{\alpha} = A$ obtida acima, usando a adjunta podemos obter o inverso desta matriz que é

$$B = [I]^{\alpha}_{\beta} = \begin{bmatrix} 3/2 & -1 & -1/2 \\ 1/2 & 0 & -1/2 \\ -1/2 & 1 & 1/2 \end{bmatrix}.$$

Portanto, calcular $[v]_{\beta} = [I]^{\alpha}_{\beta}[v]_{\alpha} = \begin{bmatrix} 3/2 & -1 & -1/2 \\ 1/2 & 0 & -1/2 \\ -1/2 & 1 & 1/2 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ -2 \end{bmatrix}.$

c) Observe que

$$\begin{aligned} T\left(\begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}\right) &= T\left(4\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix} - 2\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right) \\ &= 4\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} - 2\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ -2 \end{bmatrix}. \end{aligned}$$

□

Exercícios propostos

P7.1. Considere as bases $\alpha = \left\{\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]\right\}$, $\beta = \left\{\left[\begin{smallmatrix} 1 \\ 3 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ 4 \end{smallmatrix}\right]\right\}$ e $\gamma = \left\{\left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right], \left[\begin{smallmatrix} 2 \\ 3 \end{smallmatrix}\right]\right\}$ do $\mathbb{R}^2$. Encontre as matrizes de mudança de coordenadas nos seguintes casos: a) $[I]^{\gamma}_{\beta}$; b) $[I]^{\beta}_{\gamma}$; c) $[I]^{\beta}_{\alpha}$ e d) $[I]^{\alpha}_{\gamma}$.

P7.2. Suponha que os eixos x e y do plano $\mathbb{R}^2$ tenham sido girados 30° no sentido anti-horário para formar novos eixos x' e y' do plano. Encontre:

a) Os vetores unitários na direção dos novos eixos x' e y';

b) A matriz P de mudança de coordenadas da base antiga para a base nova;

c) As novas coordenadas dos pontos $\left[\begin{smallmatrix} 1 \\ 3 \end{smallmatrix}\right]$ e $\left[\begin{smallmatrix} 3 \\ 5 \end{smallmatrix}\right]$;

d) Por fim, verifique que $PP^t = I$.

P7.3. a) Ache a expressão da transformação linear $T : \mathbb{R}^3 \to \mathbb{R}^2$ tal que $T(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}) = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$, $T(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}) = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ e $T(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}) = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$;

b) Encontre $\mathbf{v} \in \mathbb{R}^3$ tal que $T(\mathbf{v}) = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$.

P7.4. Seja G um operador do $\mathbb{R}^2$ e α a base a seguir:

$$G\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 2x-7y \\ 4x-3y \end{bmatrix} \text{ e } \alpha = \left\{\begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} -2 \\ 5 \end{bmatrix}\right\}$$

a) Encontre a matriz $[G]_\alpha^\alpha$ de G, com respeito à α.

b) Verifique que $[G]_\alpha^\alpha [\mathbf{w}]_\alpha = [G(\mathbf{w})]_\alpha$ para o vetor $\mathbf{w} = \begin{bmatrix} 4 \\ -2 \end{bmatrix}$.

P7.5. Para cada um dos operadores lineares T do $\mathbb{R}^2$ a seguir, encontre a matriz A, que representa T (em relação à base canônica do $\mathbb{R}^2$).

a) T definida por $T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$ e $T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 5 \\ 8 \end{bmatrix}$.

b) T é a rotação no sentido anti-horário em torno da origem de $\pi/2$.

c) T é a reflexão de $\mathbb{R}^2$ em torno da reta $y = -x$.

P7.6. Mostre que a relação de semelhança entre matrizes é uma relação de equivalência, isto é, a relação é a seguinte: Dizemos que as matrizes A e B são matrizes semelhantes (e escrevemos $A \cong B$) se existe uma matriz P invertível, tal que $B = P^{-1}AP$. Mostre então que: a) $A \cong A$; b) Se $A \cong B$ então $B \cong A$ e c) Se $A \cong B$ e $B \cong C$ então $A \cong C$.

Capítulo 8

Autovalores, Autovetores e Diagonalização de Operadores

8.0 Introdução

Ao invés de fazer uma introdução tradicional, vamos, sem rodeios, considerar dois exemplos que irão motivar o assunto deste capítulo. Esses exemplos envolvem o cálculo de *potências de matrizes*. Se A é uma matriz quadrada e k é um número natural, então A^k representa a matriz $A \cdot A \cdot A \cdots A$ (k vezes). A matriz A^k chama-se a *k-ésima potência* da matriz A. Note que só faz sentido considerar potências de matrizes que forem *quadradas* (veja o exercício P8.1).

Exemplo 8.1

Calcule a k-ésima potência D^k da matriz $D = \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix}$.

Solução: Vamos primeiro calcular D^2:

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

Em seguida, calculamos D^3:

$$D^3 = D^2 \cdot D = \begin{bmatrix} 3^2 & 0 \\ 0 & (-2)^2 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix} = \begin{bmatrix} 3^2 \cdot 3 & 0 \\ 0 & (-2)^2 \cdot (-2) \end{bmatrix} = \begin{bmatrix} 3^3 & 0 \\ 0 & (-2)^3 \end{bmatrix}.$$

Não é difícil convencer-se de que, para qualquer k natural, vale a fórmula

$$D^k = \begin{bmatrix} 3^k & 0 \\ 0 & (-2)^k \end{bmatrix}. \tag{8.1}$$

Os leitores e leitoras céticos podem referir-se ao exercício resolvido R8.1. □

A simplicidade do exemplo anterior deve-se ao fato de que D é uma matriz *diagonal*, isto é, as entradas fora de sua diagonal principal são todas iguais a zero. Em geral, as potências de uma matriz não são dadas por uma fórmula tão simples como (8.1). Em particular, as potências de uma matriz *não* podem ser calculadas entrada por entrada, isto é, se $A = [a_{ij}]$, então, em geral, *não vale* $A^k = [a_{ij}^k]$.

Exemplo 8.2

Calcule a k-ésima potência A^k da matriz $A = \begin{bmatrix} 8 & 5 \\ -10 & -7 \end{bmatrix}$.

Solução: Para que não haja dúvidas quanto a isso, começamos observando que a matriz A^k *não* é igual a $\left[\begin{smallmatrix} 8^k & 5^k \\ (-10)^k & (-7)^k \end{smallmatrix}\right]$, conforme o comentário que acabamos de fazer (veja também o exercício P8.2).

Para encontrar uma fórmula para A^k, vamos utilizar um artifício que, por ora, pode parecer um *passe de mágica*. Considere as matrizes

$$P = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \qquad \text{e} \qquad D = \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix}. \tag{8.2}$$

Essas matrizes são os "coelhos na cartola", que surgem de forma aparentemente inexplicável. Sua origem será revelada no devido tempo. Por ora, repare que D é exatamente a matriz do exemplo anterior, e verifique que

$$PDP^{-1} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 8 & 5 \\ -10 & -7 \end{bmatrix} = A. \tag{8.3}$$

Ou seja, vale $A = PDP^{-1}$. Essa relação é a chave para se calcular A^k:

$$\begin{aligned} A^k &= \underbrace{AA\cdots A}_{k \text{ vezes}} = \underbrace{(PDP^{-1})(PDP^{-1})\cdots(PDP^{-1})}_{k \text{ vezes}} = \\ &= PD\underbrace{\cancel{P^{-1}P}}_{I}D\underbrace{\cancel{P^{-1}P}}_{I}D\cdots D\underbrace{\cancel{P^{-1}P}}_{I}DP^{-1} = P\underbrace{DD\cdots D}_{k \text{ vezes}}P^{-1} = PD^kP^{-1}. \end{aligned}$$

Já calculamos D^k no exemplo anterior. Aplicando o resultado lá obtido, temos

$$\begin{aligned} A^k = PD^kP^{-1} &= \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 3^k & 0 \\ 0 & (-2)^k \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 2\cdot 3^k - (-2)^k & 3^k - (-2)^k \\ -2\cdot 3^k + 2\cdot(-2)^k & -3^k + 2\cdot(-2)^k \end{bmatrix}. \end{aligned} \tag{8.4}$$

Para convencer-se de que a "mágica" funcionou, veja o exercício P8.3. □

De onde surgiu o "coelho na cartola", ou melhor, o par de matrizes P e D dado em (8.2)? O que está por trás da relação-chave $A = PDP^{-1}$?

O "truque" será revelado oportunamente, e essas perguntas serão plenamente respondidas na seção 8.4. (Até lá, pedimos paciência ao respeitável público...) Já podemos, no entanto, adiantar algumas pistas.

Sejam $\mathbf{v}_1$ e $\mathbf{v}_2$ as colunas da matriz P, isto é, $\mathbf{v}_1 = \left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right]$ e $\mathbf{v}_2 = \left[\begin{smallmatrix} -1 \\ 2 \end{smallmatrix}\right]$. Observe que a matriz A age sobre os vetores $\mathbf{v}_1$ e $\mathbf{v}_2$ de forma especial:

$$A\mathbf{v}_1 = \begin{bmatrix} 8 & 5 \\ -10 & -7 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 3 \\ -3 \end{bmatrix} = 3\begin{bmatrix} 1 \\ -1 \end{bmatrix} = 3\mathbf{v}_1 \tag{8.5}$$

$$\text{e} \quad A\mathbf{v}_2 = \begin{bmatrix} 8 & 5 \\ -10 & -7 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ -4 \end{bmatrix} = -2\begin{bmatrix} -1 \\ 2 \end{bmatrix} = -2\mathbf{v}_2. \tag{8.6}$$

Ou seja, o vetor $A\mathbf{v}_1$ é um múltiplo escalar de $\mathbf{v}_1$, e $A\mathbf{v}_2$ é um múltiplo de $\mathbf{v}_2$! Nessas circunstâncias, diremos que $\mathbf{v}_1$ e $\mathbf{v}_2$ são *autovetores* da matriz A. Os fatores de proporcionalidade 3 e -2 que aparecem nas equações acima são seus respectivos *autovalores*. Isso motiva o tema da próxima seção, onde esses conceitos serão definidos mais precisamente. Observe que os autovalores 3 e -2 são justamente as entradas na diagonal da matriz D. Isso *não* é mera coincidência!

Para concluir a introdução, observamos que o cálculo de potências de matrizes não é um exercício frívolo. Essa e outras questões correlatas são importantes, por exemplo, na resolução de *equações a diferenças lineares* e *sistemas de equações diferenciais ordinárias lineares*, que surgem em muitas aplicações.

8.1 Autovalores e autovetores

Vamos definir os conceitos de *autovalor* e *autovetor* mencionados acima.

Definição 8.3
Sejam A uma matriz $n \times n$, $\mathbf{v}$ um vetor *não-nulo* de $\mathbb{R}^n$, e λ um escalar. Se a equação vetorial

$$A\mathbf{v} = \lambda\mathbf{v} \qquad (\text{com } \mathbf{v} \neq \mathbf{0}) \tag{8.7}$$

for válida, diremos que $\mathbf{v}$ é um **autovetor** e que λ é um **autovalor** da matriz A. Diremos, mais precisamente, que $\mathbf{v}$ é um *autovetor de A associado ao autovalor* λ.

Geometricamente, a equação (8.7) diz que os vetores $A\mathbf{v}$ e $\mathbf{v}$ são colineares, com $\mathbf{v}$ não-nulo (recorde o exercício P2.19).

Exemplo 8.4
Seja A a matriz 2×2 do exemplo 8.2. As equações (8.5) e (8.6) mostram que $\mathbf{v}_1 = \left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right]$ é um autovetor de A associado ao autovalor $\lambda_1 = 3$ e que $\mathbf{v}_2 = \left[\begin{smallmatrix} -1 \\ 2 \end{smallmatrix}\right]$ é um autovetor de A associado ao autovalor $\lambda_2 = -2$.

O vetor $\mathbf{u} = \left[\begin{smallmatrix} 2 \\ -3 \end{smallmatrix}\right]$, por outro lado, *não é* um autovetor de A, pois a equação $A\mathbf{u} = \lambda\mathbf{u}$ *não* é válida para valor algum do escalar λ. De fato,

$$A\mathbf{u} = \begin{bmatrix} 8 & 5 \\ -10 & -7 \end{bmatrix} \begin{bmatrix} 2 \\ -3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \neq \lambda \begin{bmatrix} 2 \\ -3 \end{bmatrix} = \lambda\mathbf{u},$$

isto é, a equação $\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right] = \lambda\left[\begin{smallmatrix} 2 \\ -3 \end{smallmatrix}\right]$ *não* é satisfeita para valor algum de λ.

USAR O EXEMPLO ANTERIOR PARA MOTIVAR A DISCUSSAO GEOMETRICA DE AUTOVALORES E AUTOVETORES DE OPERADORES. !!! ALGO ASSIM: Pensando em A como um operador linear em $\mathbb{R}^2$, o efeito de A sobre $\mathbf{v}_1$ é simplesmente uma "dilatação" pelo fator 3. O efeito sobre $\mathbf{v}_2$ é de dilatação pelo fator 2, seguida de uma reflexão com respeito à origem (ou seja, uma "reversão" no sentido do vetor, mas preservando sua direção). Por outro lado, o operador A *não* preserva a direção do vetor $\mathbf{u}$, isto é, o vetor $A\mathbf{u}$ *não* é colinear a $\mathbf{u}$.

COLOCAR UMA FIGURA mostrando as ações de A sobre $\mathbf{v}_1$, $\mathbf{v}_2$ e $\mathbf{u}$.

Exemplo 8.5
Seja A a matriz 2×2 do exemplo 8.2. Encontre todos os autovetores de A associados ao autovalor $\lambda_1 = 3$.

Solução: Queremos achar todas as soluções de $A\mathbf{x} = 3\mathbf{x}$, com $\mathbf{x} \neq \mathbf{0}$. A equação vetorial $A\mathbf{x} = 3\mathbf{x}$ é simplesmente um sistema linear. De fato, temos

$$A\mathbf{x} = 3\mathbf{x} \;\Leftrightarrow\; A\mathbf{x} - 3\mathbf{x} = \mathbf{0} \;\Leftrightarrow\; A\mathbf{x} - 3I_2\mathbf{x} = \mathbf{0} \;\Leftrightarrow\; (A - 3I_2)\mathbf{x} = \mathbf{0}, \qquad (8.8)$$

onde I_2 representa a matriz identidade 2×2, de forma que $I_2\mathbf{x} = \mathbf{x}$ para qualquer vetor $\mathbf{x} \in \mathbb{R}^2$. Estamos buscando, portanto, as soluções *não-triviais* do sistema homogêneo $(A - 3I_2)\mathbf{x} = \mathbf{0}$, cuja matriz de coeficientes $A - 3I_2$ é dada por

$$A - 3I_2 = \begin{bmatrix} 8 & 5 \\ -10 & -7 \end{bmatrix} - 3\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 8-3 & 5 \\ -10 & -7-3 \end{bmatrix} = \begin{bmatrix} 5 & 5 \\ -10 & -10 \end{bmatrix}.$$

Para resolver o sistema, vamos escalonar sua matriz completa $\left[A - 3I_2 \mid \mathbf{0}\right]$:

$$\left[\begin{array}{cc|c} 5 & 5 & 0 \\ -10 & -10 & 0 \end{array}\right] \xrightarrow{\ell_2 \to \ell_2 + 2\ell_1} \left[\begin{array}{cc|c} 5 & 5 & 0 \\ 0 & 0 & 0 \end{array}\right] \xrightarrow{\ell_1 \to \frac{1}{5}\ell_1} \left[\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right].$$

Assim, o sistema $(A - 3I_2)\mathbf{x} = \mathbf{0}$ é equivalente a $x_1 + x_2 = 0$. A descrição vetorial paramétrica de seu conjunto-solução é

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix}. \qquad (8.9)$$

Isso mostra que os autovetores de A associados a $\lambda_1 = 3$ são os múltiplos *não-nulos* do vetor $\mathbf{u}_1 = \left[\begin{smallmatrix} -1 \\ 1 \end{smallmatrix}\right]$, ou seja, são os vetores da forma (8.9), com $x_2 \neq 0$. Observe que o autovetor $\mathbf{v}_1 = \left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right]$ do exemplo 8.4 corresponde a $x_2 = -1$. $\square$

Observação 8.6
A equação $A\mathbf{v} = \lambda\mathbf{v}$ da definição 8.3 é equivalente a $(A - \lambda I_n)\mathbf{v} = \mathbf{0}_n$, onde I_n é a matriz identidade $n \times n$ e $\mathbf{0}_n$ é o vetor zero de $\mathbb{R}^n$. Para provar a equivalência, basta generalizar o argumento feito em (8.8):

$$A\mathbf{v} = \lambda\mathbf{v} \;\Leftrightarrow\; A\mathbf{v} - \lambda\mathbf{v} = \mathbf{0}_n \;\Leftrightarrow\; A\mathbf{v} - \lambda I_n\mathbf{v} = \mathbf{0}_n \;\Leftrightarrow\; (A - \lambda I_n)\mathbf{v} = \mathbf{0}_n.$$

Esse fato, apesar de simples, será utilíssimo. Note que, se *fixarmos* um escalar λ, e pensarmos no vetor $\mathbf{v}$ como a "incógnita", então $(A - \lambda I)\mathbf{v} = \mathbf{0}$ torna-se nada mais do que um sistema linear homogêneo. Em particular, se λ for um autovalor de A, então os autovetores de A associados a λ serão precisamente as soluções *não-triviais* desse sistema.

Na definição 8.3, introduzimos simultaneamente os conceitos de autovetor e de autovalor. Essa abordagem é bastante clara (ou assim esperamos!), mas é importante compreender a seguinte reformulação, que é um pouco mais sutil.

Definição 8.7 (Definição 8.3 reformulada)
Seja A uma matriz $n \times n$. Diremos que um vetor *não-nulo* $\mathbf{v} \in \mathbb{R}^n$ é um **autovetor** de A quando existir um escalar λ tal que $A\mathbf{v} = \lambda\mathbf{v}$. Similarmente, diremos que um escalar λ é um **autovalor** de A quando existir um vetor $\mathbf{v} \neq \mathbf{0}$ tal que $A\mathbf{v} = \lambda\mathbf{v}$.

Exemplo 8.8
Determine se $\mu = 2$ é um autovalor da matriz A do exemplo 8.2.

Solução: Conforme a definição 8.7, $\mu = 2$ será um autovalor de A se existir um vetor *não-nulo* $\mathbf{v}$ tal que $A\mathbf{v} = 2\mathbf{v}$. Pela observação 8.6, essa equação é equivalente a $(A - 2I)\mathbf{v} = \mathbf{0}$. Assim, 2 será um autovalor de A se, e somente se, esse sistema tiver soluções *não-triviais*. Sua matriz de coeficientes é dada por

$$A - 2I = \begin{bmatrix} 8 & 5 \\ -10 & -7 \end{bmatrix} - 2\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 8-2 & 5 \\ -10 & -7-2 \end{bmatrix} = \begin{bmatrix} 6 & 5 \\ -10 & -9 \end{bmatrix}.$$

Você deve verificar que ambas as colunas dessa matriz são colunas-pivô. Pela proposição 3.3, o sistema $(A - 2I)\mathbf{v} = \mathbf{0}$ possui unicamente a solução trivial $\mathbf{v} = \mathbf{0}$. Dessa maneira, $\mu = 2$ *não* é um autovalor de A. □

A proposição a seguir generaliza esse exemplo e será crucial na próxima seção.

Proposição 8.9
Seja A uma matriz $n \times n$. Um escalar λ será um autovalor de A se, e somente se, a matriz $A - \lambda I$ não *for invertível.*

Demonstração: Por definição, o escalar λ será um autovalor de A quando existir um vetor *não-nulo* $\mathbf{v}$ tal que $A\mathbf{v} = \lambda\mathbf{v}$. Pela observação 8.6, essa equação é equivalente a $(A - \lambda I)\mathbf{v} = \mathbf{0}$. Assim, λ será um autovalor de A quando o sistema homogêneo $(A - \lambda I)\mathbf{v} = \mathbf{0}$ tiver soluções *não-triviais*. De acordo com a teoria do capítulo 5, isso ocorrerá se, e somente se, a matriz $A - \lambda I$ *não* for invertível. □

O corolário a seguir é obtido ao considerar-se o caso $\lambda = 0$ nesta proposição.

Corolário 8.10
Seja A uma matriz $n \times n$. O escalar zero será um autovalor de A se, e somente se, a matriz A for não-invertível.

Enfatizamos que, por definição, um *autovetor* de A é necessariamente um vetor *não-nulo* (ver ???). O *escalar zero*, por outro lado, pode ser *autovalor* de uma matriz, conforme o corolário acima.

Observação 8.11
Note que só faz sentido definir autovalores e autovetores de matrizes *quadradas*. De fato, suponha que A seja uma matriz $n \times m$, e considere a equação $A\mathbf{v} = \lambda\mathbf{v}$. Para que o produto $A\mathbf{v}$ esteja definido, é necessário que o vetor $\mathbf{v}$ pertença a $\mathbb{R}^m$. Sendo assim, a equação $A\mathbf{v} = \lambda\mathbf{v}$ representa a igualdade entre o vetor $A\mathbf{v}$ de $\mathbb{R}^n$ e o vetor $\lambda\mathbf{v}$ de $\mathbb{R}^m$, o que só faz sentido se n for igual a m.

8.2 O polinômio característico

Para cada escalar λ *fixo*, a equação $A\mathbf{v} = \lambda\mathbf{v}$ equivale ao sistema linear homogêneo $(A - \lambda I)\mathbf{v} = \mathbf{0}$, como verificamos na observação 8.6 (recorde, também, o exemplo 8.5). Assim, se os autovalores da matriz A forem conhecidos, podemos facilmente encontrar seus autovetores.

Não é tão fácil, no entanto, determinar simultaneamente os autovalores e autovetores de uma matriz (veja o exercício P8.4). Como devemos proceder, então, para determinar os autovalores de A, ou, ao menos, caracterizá-los de forma simples?

Já estamos equipados para responder essa questão. Aliando o teorema 6.25 à proposição 8.9, obtemos o resultado principal desta seção:

Teorema 8.12
Seja A uma matriz $n \times n$. Um escalar λ será um autovalor de A se, e somente se, satisfizer a equação $\det(A - \lambda I) = 0$.

A equação $\det(A - xI) = 0$, onde x é uma "incógnita" escalar, é chamada a **equação característica** da matriz A.[1] O teorema acima diz que os autovalores de uma matriz são exatamente as soluções de sua equação característica.

Exemplo 8.13
Encontre todos os autovalores da matriz $A = \begin{bmatrix} 8 & 5 \\ -10 & -7 \end{bmatrix}$ do exemplo 8.2.

Solução: Vamos calcular $\det(A - xI)$:

$$\det(A - xI) = \det\left(\begin{bmatrix} 8 & 5 \\ -10 & -7 \end{bmatrix} - x\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) = \det\begin{bmatrix} 8 - x & 5 \\ -10 & -7 - x \end{bmatrix} =$$

$$= (8 - x)\cdot(-7 - x) - 5\cdot(-10) = x^2 - x - 6.$$

Pelo teorema 8.12, os autovalores de A são as soluções da equação $x^2 - x - 6 = 0$. O polinômio $x^2 - x - 6$ é fatorado na forma $(x - 3)(x + 2)$, logo suas raízes são $\lambda_1 = 3$ e $\lambda_2 = -2$ (pode-se também usar a "fórmula de Báskara"). Já sabíamos, em vista das equações (8.5) e (8.6), que $\lambda_1 = 3$ e $\lambda_2 = -2$ eram autovalores de A. Agora, sabemos não há outros: λ_1 e λ_2 são os *únicos* autovalores da matriz A. □

Observe, no exemplo acima, que $\det(A - xI)$ é um polinômio de grau 2 em x. Pode-se demonstrar que, se A for uma matriz $n \times n$, então $\det(A - xI)$ será um polinômio de grau n na variável x. Este é chamado o **polinômio característico** de A. Denota-se o polinômio característico de A por $p_A(x)$.

Nessa terminologia, o teorema 8.12 diz que *os autovalores de uma matriz A são dados pelas raízes de seu polinômio característico.* A **multiplicidade algébrica** de um autovalor λ de A é definida como a multiplicidade de λ como raiz do polinômio característico.

[1] Lembre-se de que o determinante de uma matriz quadrada é um *escalar*. Dessa maneira, a equação característica é uma equação escalar, e não vetorial ou matricial.

Exemplo 8.14
Obtenha o polinômio característico da matriz

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

Em seguida, determine os autovalores de B e suas respectivas multiplicidades.

Solução: A matriz $B - xI$ é dada por

$$\begin{bmatrix} 7 & 1 & -4 & 0 \\ 0 & -1 & 2 & 3 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 7 \end{bmatrix} - \begin{bmatrix} x & 0 & 0 & 0 \\ 0 & x & 0 & 0 \\ 0 & 0 & x & 0 \\ 0 & 0 & 0 & x \end{bmatrix} = \begin{bmatrix} 7-x & 1 & -4 & 0 \\ 0 & -1-x & 2 & 3 \\ 0 & 0 & -x & 1 \\ 0 & 0 & 0 & 7-x \end{bmatrix}.$$

Como essa é uma matriz triangular, seu determinante é simplesmente o produto dos elementos na diagonal (isso é um caso particular do teorema 6.23). O polinômio característico de B, portanto, é

$$p_B(x) = \det(B - xI) = (7-x)(-1-x)(-x)(7-x) = -x(7-x)^2(-1-x).$$

Os autovalores de B são dados pelas raízes de p_B, ou seja, pelas soluções da equação $p_B(x) = 0$. Examinando a expressão de p_B, acima, fica evidente que os autovalores de B são:

$$\begin{aligned} \lambda_1 &= 0, && \text{com multiplicidade igual a 1;} \\ \lambda_2 &= 7, && \text{com multiplicidade igual a 2; e} \\ \lambda_3 &= -1, && \text{com multiplicidade igual a 1.} \end{aligned}$$

□

Observe que os autovalores da matriz B do exemplo acima são justamente as entradas na diagonal de B. Isso se deve ao fato de que B é uma matriz triangular. É fácil generalizar esse exemplo para demonstrar o seguinte resultado.

Proposição 8.15
Se A é uma matriz $n \times n$ triangular (superior ou inferior), então seus autovalores são dados pelas entradas em sua diagonal principal.

Essa proposição vale, em particular, se A é uma matriz *diagonal* (pois, nesse caso, A é triangular superior e inferior, simultaneamente).

Nem toda matriz possui autovalores reais, como mostra o exemplo a seguir.

Exemplo 8.16
Obtenha as raízes do polinômio característico da matriz $C = \begin{bmatrix} 3 & -2 \\ 2 & 3 \end{bmatrix}$.

Solução: O polinômio característico de C é dado por

$$p_C(x) = \det(C - xI) = \det \begin{bmatrix} 3-x & -2 \\ 2 & 3-x \end{bmatrix} = (3-x)(3-x)+4 = x^2 - 6x + 13.$$

Para achar as raízes de $p_C(x)$, usamos a fórmula de Báskara:

$$\lambda = \frac{6 \pm \sqrt{6^2 - 4 \cdot 13}}{2} = \frac{6 \pm \sqrt{-16}}{2} = \frac{6 \pm 4\sqrt{-1}}{2} = 3 \pm 2\sqrt{-1}.$$

Assim, as raízes de $p_C(x)$ são os números complexos conjugados $\lambda_1 = 3 + 2i$ e $\lambda_2 = 3 - 2i$, onde i representa a unidade imaginária. Na seção **??**, estudaremos mais a fundo matrizes com “autovalores complexos”. Por ora, basta constatar que a matriz C não tem autovalores reais. A matriz C tampouco possui autovetores em $\mathbb{R}^2$. De fato, se C possuísse um autovetor $\mathbf{v} \in \mathbb{R}^2$, então teria que existir algum autovalor (real) associado a $\mathbf{v}$. □

Recorde que o polinômio característico de uma matriz $n \times n$ é um polinômio de grau n, e que suas raízes são os autovalores da matriz em questão. Assim, uma matriz de tamanho $n \times n$ terá exatamente n autovalores, *se levarmos em conta os os autovalores complexos e as multiplicidades.*[2] Entretanto, é claro que uma matriz $n \times n$ pode ter menos do que n autovalores *reais* e *distintos*. A matriz 4×4 do exemplo 8.14 tem apenas três autovalores reais distintos (um deles tem multiplicidade igual a dois) e a matriz 2×2 do exemplo 8.16 não possui autovalor real algum.

Achar as raízes de um polinômio de grau n não é uma tarefa fácil, em geral. O caso $n = 1$ é muito simples, e temos a fórmula de Báskara para o caso $n = 2$. Existem ainda fórmulas para os casos $n = 3$ e $n = 4$, mas elas não são tão simples. Em contrapartida, não existem fórmulas gerais para determinar as raízes de polinômios de grau maior que ou igual a cinco.[3]

Na prática, os autovalores de matrizes grandes são determinados aproximadamente, por meio de métodos computacionais.[4] Em certos casos especiais, no entanto, é possível determinar com exatidão os autovalores de uma matriz de tamanho arbitrário. Um exemplo é o caso das matrizes triangulares, conforme a proposição 8.15.

O teorema a seguir será importante na seção 8.4. Recorde a definição 7.11: duas matrizes quadradas A e B são ditas *semelhantes* se existe uma matriz invertível P tal que $B = P^{-1}AP$.

Teorema 8.17
Se A e B forem matrizes $n \times n$ semelhantes, então elas terão o mesmo polinômio característico. Consequentemente, matrizes semelhantes A e B terão os mesmos autovalores, com as mesmas multiplicidades.

[2]Isso decorre do *teorema fundamental da álgebra*, que diz que todo polinômio de grau n tem exatamente n raízes, se levarmos em conta as raízes complexas e as multiplicidades.

[3]Não se trata, meramente, de que essas fórmulas sejam ainda desconhecidas. O matemático norueguês Niels Abel demonstrou, em 1824, que tais fórmulas, de fato, não existem!

[4]Diversos bons métodos existem, muitos dos quais calculam os autovalores e os autovetores simultaneamente. Esse assunto, no entanto, foge ao escopo deste texto.

Demonstração: Seja P uma matriz invertível tal que $B = P^{-1}AP$. Então

$$B - xI = \underbrace{P^{-1}AP}_{B} - x\underbrace{P^{-1}IP}_{I} = P^{-1}(A - xI)P.$$

Dessa maneira, usando os teoremas 6.24 e 6.25, obtemos

$$\begin{aligned} p_B(x) &= \det(B - xI) = \det\big(P^{-1}(A - xI)P\big) = \\ &= \det(P^{-1}) \cdot \det(A - xI) \cdot \det(P) = \\ &= \underbrace{\det(P^{-1}) \cdot \det(P)}_{1} \cdot \det(A - xI) = \det(A - xI) = p_A(x). \end{aligned}$$ □

8.3 Autoespaços

Na seção anterior, obtivemos uma caracterização dos autovalores de uma dada matriz. Cada autovalor está associado a um certo conjunto de autovetores. Nesta seção, vamos estudar algumas propriedades de tais conjuntos. MUDAR?: ESSES CONJUNTOS EM QUE OS AUTOVETORES "MORAM" TÊM PROPRIEDADES INTERESSANTES. O OBJETIVO DESTA SEÇÃO É ESTUDÁ-LOS. ???

Como vimos na observação 8.6, o conjunto dos autovetores de uma matriz quadrada A associados a um autovalor λ coincide com o conjunto das soluções *não-triviais* do sistema $(A - \lambda I)\mathbf{x} = \mathbf{0}$. O conjunto de *todas* as soluções desse sistema linear homogêneo é simplesmente o núcleo da matriz $A - \lambda I$ (veja a definição 3.8). Isso motiva a definição a seguir.

Definição 8.18
Seja λ um autovalor de uma matriz quadrada A. O núcleo da matriz $A - \lambda I$ é chamado de **autoespaço** de A associado a λ, e é denotado por $E(A, \lambda)$.

Podemos sintetizar essa discussão, em símbolos, da seguinte maneira:

$$E(A, \lambda) = \mathrm{Nuc}(A - \lambda I) = \{\mathbf{0}\} \cup \{\text{autovetores de } A \text{ associados a } \lambda\}.$$

O símbolo "∪" representa a união de conjuntos. A equação acima diz que o autoespaço $E(A, \lambda)$ consiste do conjunto de todos os autovetores de A associados a λ, acrescido do vetor zero (que, conforme a definição 8.3, *não* é um autovetor).

Pela proposição 4.7, um autoespaço de uma matriz $n \times n$ é um *subespaço* de $\mathbb{R}^n$.

MELHORAR A DISCUSSÃO ACIMA. ENFATIZAR: AUTOESPAÇOS SÃO SUBESPAÇOS, E OS AUTOVETORES "MORAM" NESSES SUBESPAÇOS. (NÃO EXISTEM AUTOVETORES "PERDIDOS/ISOLADOS".)

DIZER QUE UM OPERADOR A AGE DE FORMA "MUITO SIMPLES" SOBRE SEUS AUTOESPAÇOS E COLOCAR UMA FIGURA. MENCIONAR QUE AUTOESPAÇOS SÃO *INVARIANTES* SOB A AÇÃO DE A.

Nos exemplos a seguir, consideramos questões que serão importantes na seção 8.4: encontrar uma base e determinar a dimensão de cada autoespaço de uma dada matriz.

Exemplo 8.19
Encontre uma base de cada autoespaço da matriz $A = \left[\begin{smallmatrix} 8 & 5 \\ -10 & -7 \end{smallmatrix}\right]$ do exemplo 8.2. Em seguida, determine a dimensão de cada um.

Solução: Os autovalores de A são $\lambda_1 = 3$ e $\lambda_2 = -2$, como vimos no exemplo 8.13. Assim, a matriz A tem dois autoespaços: $E(A, \lambda_1)$ e $E(A, \lambda_2)$.

Vamos, primeiro, encontrar uma base de $E(A, \lambda_1)$, ou seja, de $\operatorname{Nuc}(A - \lambda_1 I) = \operatorname{Nuc}(A - 3I)$. Para isso, usamos o procedimento sugerido na subseção 4.5.1.[5] Queremos resolver o sistema $(A - 3I)\mathbf{x} = \mathbf{0}$. Essa tarefa já foi realizada no exemplo 8.5, em que obtivemos a descrição (8.9) do conjunto-solução. Assim, o conjunto $\{\mathbf{u}_1\} = \left\{\left[\begin{smallmatrix} -1 \\ 1 \end{smallmatrix}\right]\right\}$ é uma base de $E(A, \lambda_1) = \operatorname{Nuc}(A - \lambda_1 I)$. De fato, a descrição (8.9) mostra que o conjunto $\{\mathbf{u}_1\}$ gera $E(A, \lambda_1)$, e é claro que esse conjunto é linearmente independente.

Agora, vamos obter uma base de $E(A, \lambda_2) = \operatorname{Nuc}(A - \lambda_2 I) = \operatorname{Nuc}(A + 2I)$. Seguindo, mais uma vez, o procedimento da subseção 4.5.1, vamos resolver o sistema $(A + 2I)\mathbf{x} = \mathbf{0}$, por escalonamento de sua matriz completa:

$$\begin{bmatrix} A + 2I \mid \mathbf{0} \end{bmatrix} = \left[\begin{array}{cc|c} 8+2 & 5 & 0 \\ -10 & -7+2 & 0 \end{array}\right] = \left[\begin{array}{cc|c} 10 & 5 & 0 \\ -10 & -5 & 0 \end{array}\right] \longrightarrow$$

$$\xrightarrow{\ell_2 \to \ell_2 + \ell_1} \left[\begin{array}{cc|c} 10 & 5 & 0 \\ 0 & 0 & 0 \end{array}\right] \xrightarrow{\ell_1 \to \frac{1}{10}\ell_1} \left[\begin{array}{cc|c} 1 & 1/2 & 0 \\ 0 & 0 & 0 \end{array}\right].$$

Assim, o sistema $(A + 2I)\mathbf{x} = \mathbf{0}$ equivale à equação $x_1 + \frac{1}{2}x_2 = 0$, e seu conjunto-solução é descrito por

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} -1/2 \\ 1 \end{bmatrix}. \tag{8.10}$$

Isso mostra que o conjunto $\{\mathbf{u}_2\} = \left\{\left[\begin{smallmatrix} -1/2 \\ 1 \end{smallmatrix}\right]\right\}$ é uma base de $E(A, \lambda_2)$. Observe que o autovetor $\mathbf{v}_2 = \left[\begin{smallmatrix} -1 \\ 2 \end{smallmatrix}\right]$ do exemplo 8.4 é igual a $2\mathbf{u}_2$, ou seja, $\mathbf{v}_2$ corresponde à escolha $x_2 = 2$ na equação (8.10).

Resta apenas determinar as dimensões de $E(A, \lambda_1)$ e de $E(A, \lambda_2)$. Ora, cada um desses subespaços possui uma base contendo apenas *um* vetor, portanto cada um deles tem dimensão igual a 1. □

Exemplo 8.20
Seja

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

Sabendo que $\lambda = 5$ é um autovalor de A, encontre uma base do autoespaço $E(A, \lambda)$ e determine a sua dimensão.

[5]Sugerimos, neste momento, uma rápida revisão dessa subseção.

Solução: Vamos resolver o sistema homogêneo $(A - \lambda I)\mathbf{x} = \mathbf{0}$, onde $\lambda = 5$. Para isso, escalonamos sua matriz completa:

$$\big[A - \lambda I \mid \mathbf{0}\big] = \left[\begin{array}{ccc|c} 4-5 & 1 & -2 & 0 \\ 0 & 5-5 & 0 & 0 \\ -1 & 1 & 3-5 & 0 \end{array}\right] = \left[\begin{array}{ccc|c} -1 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 1 & -2 & 0 \end{array}\right] \longrightarrow$$

$$\xrightarrow{\ell_1 \to -\ell_1} \left[\begin{array}{ccc|c} 1 & -1 & 2 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 1 & -2 & 0 \end{array}\right] \xrightarrow{\ell_3 \to \ell_3 - \ell_1} \left[\begin{array}{ccc|c} 1 & -1 & 2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

O sistema proposto equivale, portanto, à equação $x_1 - x_2 + 2x_3 = 0$ (as variáveis livres são x_2 e x_3). O conjunto-solução é descrito por

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_2 - 2x_3 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \underbrace{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}_{\mathbf{v}_1} + x_3 \underbrace{\begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}}_{\mathbf{v}_2}. \tag{8.11}$$

Pela teoria da subseção 4.5.1, os vetores $\mathbf{v}_1$ e $\mathbf{v}_2$ indicados acima compõem uma base de $E(A, \lambda) = \mathrm{Nuc}(A - 5I)$. Como a base $\{\mathbf{v}_1, \mathbf{v}_2\}$ possui *dois* elementos, a dimensão de $E(A, \lambda)$ é igual a 2. Em símbolos: $\dim E(A, \lambda) = 2$. □

Seja λ um autovalor de uma matriz quadrada A. Recorde que a *multiplicidade algébrica* de λ é a sua multiplicidade enquanto raiz do polinômio característico de A. A **multiplicidade geométrica** de λ é definida como a dimensão do autoespaço de A associado a λ. Em outras palavras, a multiplicidade geométrica do autovalor λ é o número $\dim E(A, \lambda)$.

Teorema 8.21
Seja A uma matriz quadrada. Para cada autovalor λ de A, a multiplicidade geométrica de λ é menor que ou igual à sua multiplicidade algébrica.

A demonstração desse teorema pode parecer um pouco abstrata, portanto é relegada à seção 8.7.

A FAZER:

- Botar um exemplo em que vale a desigualdade estrita: Na matriz B do exemplo 8.14, o autovalor $\lambda_2 = 7$ tem multiplicidade algébrica igual a 2, mas multiplicidade geométrica igual a 1.
- Mencionar tb que $1 \leqslant \dim E(A, \lambda)$. Sintetizar a discussão numa observação que diz $1 \leqslant \dim E(A, \lambda) \leqslant n_\lambda$.
- Dizer algo "to the effect" que, agora, estudaremos um pouco a relação entre autoespaços distintos. Eles são, em certo sentido, independentes.

Teorema 8.22
Sejam $\mathbf{v}_1$, $\mathbf{v}_2$, ... , $\mathbf{v}_m$ autovetores de uma matriz A associados, respectivamente, a autovalores distintos λ_1, λ_2, ... , λ_m. *Nessas condições, o conjunto $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ é linearmente independente.*

Resumidamente, o teorema acima diz que autovetores associados a autovalores distintos são linearmente independentes. O resultado a seguir é simultaneamente uma consequência e uma generalização do teorema.

Corolário 8.23
Sejam $\lambda_1,\ \lambda_2,\ \dots\ ,\ \lambda_m$ *autovalores* distintos *de uma matriz* A *e sejam* $\mathcal{B}_1,\ \mathcal{B}_2,\ \dots\ ,$ $\mathcal{B}_m$ *bases dos autoespaços* $E(A,\lambda_1),\ E(A,\lambda_2),\ \dots\ ,\ E(A,\lambda_m)$*, respectivamente. Então, o conjunto* $\mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_m$ *é linearmente independente.*

As provas desses resultados também são relegadas à seção 8.7. Vamos, no entanto, pormenorizar o enunciado do corolário, pois ele será útil na seção 8.4.

Para cada $k = 1,\ 2,\ \dots\ ,\ m$, vamos denotar os vetores de $\mathcal{B}_k$ por $\mathbf{v}_1^k,\ \dots\ ,\ \mathbf{v}_{p_k}^k$. Por hipótese, $\mathcal{B}_k$ é uma base de $E(A,\lambda_k)$.[6] Assim, cada $\mathbf{v}_j^k$ é um vetor não-nulo de $E(A,\lambda_k)$, ou seja, é um autovetor de A associado a λ_k. Além disso, cada um dos conjuntos $\mathcal{B}_k = \{\mathbf{v}_1^k, \dots, \mathbf{v}_{p_k}^k\}$ é linearmente independente. O corolário 8.23 diz que a *união* desses conjuntos,

$$\mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_m =$$
$$= \Big\{ \underbrace{\mathbf{v}_1^1, \mathbf{v}_2^1, \dots, \mathbf{v}_{p_1}^1}_{\text{vetores de } \mathcal{B}_1},\ \underbrace{\mathbf{v}_1^2, \mathbf{v}_2^2, \dots, \mathbf{v}_{p_2}^2}_{\text{vetores de } \mathcal{B}_2},\ \dots\ ,\ \underbrace{\mathbf{v}_1^m, \mathbf{v}_2^m, \dots, \mathbf{v}_{p_m}^m}_{\text{vetores de } \mathcal{B}_m} \Big\}, \tag{8.12}$$

é, ainda, linearmente independente, desde que os autovalores λ_k sejam distintos.

- A FAZER: Em um aparte, dizer que autoespaços distintos estão em soma direta/são independentes (terminologia do Elon/do Halmos ??? (conferir)).

8.4 Diagonalização de operadores

Nesta seção, iremos, finalmente, revelar o "truque" empregado no exemplo 8.2 da introdução. Recomendamos que o exemplo seja relido neste momento, pois ele motiva a definição a seguir.

Definição 8.24
Uma matriz quadrada A é dita **diagonalizável** quando ela é semelhante[7] a uma matriz diagonal, isto é, quando existem uma matriz invertível P e uma matriz diagonal D tais que $A = PDP^{-1}$ ou, equivalentemente, $D = P^{-1}AP$.

Uma fatoração de A na forma $A = PDP^{-1}$, onde P é invertível e D é diagonal, quando existe, chama-se **diagonalização** da matriz A.

Por abuso de linguagem, às vezes nos referimos, também, à "diagonalização do *operador* A", interpretando a matriz A como o operador linear $\mathbf{x} \mapsto A\mathbf{x}$ em $\mathbb{R}^n$.

A equivalência entre as relações $A = PDP^{-1}$ e $D = P^{-1}AP$, mencionada na definição acima, é/foi??? estabelecida no exercício resolvido R8.2.

[6] Repare, então, que o número p_k de vetores em $\mathcal{B}_k$ denota a dimensão do autoespaço $E(A,\lambda_k)$, ou seja, p_k é a *multiplicidade geométrica* de λ_k.

[7] A palavra "semelhante", nesse contexto, tem um significado matemático preciso. Recorde a definição 7.11.

A equação (8.3) mostra que a matriz A do exemplo 8.2 é diagonalizável. Note que a matriz D dada em (8.2) é diagonal e P é invertível. Nem toda matriz é diagonalizável, no entanto, como mostrarão os exemplos **??**.

FALAR DA IMPORTÂNCIA DA DIAGONALIZAÇÃO. COMENTAR QUE OS OPERADORES DIAGONALIZÁVEIS SÃO, EM CERTO SENTIDO, OS MAIS SIMPLES. ???

As metas desta seção são caracterizar as matrizes diagonalizáveis e estudar o processo de diagonalização. Os resultados fundamentais são os seguintes:

Teorema 8.25
Uma matriz A de tamanho $n \times n$ será diagonalizável se, e somente se, existir uma base de $\mathbb{R}^n$ composta por autovetores de A.

Com efeito, pode-se dizer mais:

Teorema 8.25′
Seja A uma matriz $n \times n$. As seguintes afirmativas são equivalentes:

(a) *Vale a relação de semelhança $A = PDP^{-1}$, onde*

$$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix} \tag{8.13}$$

é uma matriz $n \times n$ invertível e

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}, \tag{8.14}$$

é uma matriz $n \times n$ diagonal.

(b) *O conjunto $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ é uma base de $\mathbb{R}^n$ composta por autovetores da matriz A associados, respectivamente, aos autovalores λ_1, λ_2, ... , λ_n.*

Os teoremas acima são, efetivamente, duas formulações do *mesmo resultado*. Perceba que a afirmativa (a) do teorema 8.25′ diz que a matriz A é diagonalizável, conforme a definição 8.24. A afirmativa (b), por sua vez, diz que existe uma base de $\mathbb{R}^n$ (indicada, no teorema, por $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$) composta por autovetores de A. Assim, o teorema 8.25 é, essencialmente, uma *síntese* do teorema 8.25′, e segue deste como consequência imediata. O exercício **??** propõe uma reflexão mais detalhada sobre esse argumento.

Cuidado com a linguagem matemática: o teorema 8.25′ *não* diz que as afirmativas (a) e (b) são verdadeiras para *qualquer* matriz A! Ele diz apenas que, *se* (a) for verdadeira para uma determinada matriz A, *então* (b) também o será, e vice-versa. Para matrizes que *não* forem diagonalizáveis, ambas as afirmativas serão *falsas* (veja os exemplos ????).

Observe, também, que os escalares λ_1, λ_2, ... , λ_n indicados no enunciado *não* são, necessariamente, todos distintos (veja a observação 8.26, logo adiante).

Demonstração dos teoremas 8.25 e 8.25′: Como indicamos acima, o teorema 8.25 é uma consequência imediata do teorema 8.25′. Basta demonstrar, portanto, este segundo teorema.

Antes de abordar a equivalência entre as afirmativas (a) e (b), faremos algumas observações preliminares. Se P for uma matriz $n \times n$ *qualquer* (não necessariamente invertível) com colunas dadas por $\mathbf{v}_1, \ldots, \mathbf{v}_n$, então

$$AP = A\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\end{bmatrix} = \begin{bmatrix}A\mathbf{v}_1 & A\mathbf{v}_2 & \cdots & A\mathbf{v}_n\end{bmatrix}. \tag{8.15}$$

Além disso, se D for *qualquer* matriz diagonal, com elementos na diagonal principal dados por $\lambda_1, \ldots, \lambda_n$, então

$$PD = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\end{bmatrix}\begin{bmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \lambda_n\end{bmatrix} = \begin{bmatrix}\lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \cdots & \lambda_n\mathbf{v}_n\end{bmatrix}. \tag{8.16}$$

Vamos, agora, provar que a afirmativa (a) implica (b). Por hipótese, vale $A = PDP^{-1}$. Multiplicando ambos os lados desta igualdade, à direita, por P, obtemos $AP = PD$. Usando (8.15) e (8.16), esta relação se escreve na forma

$$\begin{bmatrix}A\mathbf{v}_1 & A\mathbf{v}_2 & \cdots & A\mathbf{v}_n\end{bmatrix} = \begin{bmatrix}\lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \cdots & \lambda_n\mathbf{v}_n\end{bmatrix}. \tag{8.17}$$

Esta igualdade matricial equivale às n igualdades entre as respectivas colunas:

$$A\mathbf{v}_1 = \lambda_1\mathbf{v}_1, \quad A\mathbf{v}_2 = \lambda_2\mathbf{v}_2, \quad \ldots \quad , \quad A\mathbf{v}_n = \lambda_n\mathbf{v}_n. \tag{8.18}$$

Como, por hipótese, a matriz P é invertível, os seus vetores-coluna $\mathbf{v}_1, \ldots, \mathbf{v}_n$ compõem uma base de $\mathbb{R}^n$. Em particular, esses vetores são linearmente independentes e, portanto, são todos não-nulos (veja o exercício resolvido R3.1). Sendo assim, as relações em (8.18) mostram que $\mathbf{v}_1, \ldots, \mathbf{v}_n$ são autovetores de A, associados a $\lambda_1, \ldots, \lambda_n$, respectivamente. Isso termina a prova de que (a) implica (b).

Para mostrar que (b) implica (a), vamos, essencialmente, refazer esse argumento, de trás para frente. Por hipótese, agora, $\mathbf{v}_1, \ldots, \mathbf{v}_n$ são autovetores da matriz A associados a $\lambda_1, \ldots, \lambda_n$. Sendo assim, desta vez, *começamos* com as relações (8.18), que equivalem à equação matricial (8.17). Em seguida, definimos, ou "construímos", as matrizes P e D de acordo com as equações (8.13) e (8.14). Combinando, então, as equações (8.15), (8.16) e (8.17), obtemos $AP = PD$. Por hipótese, os vetores $\mathbf{v}_1, \ldots, \mathbf{v}_n$ compõem uma base de $\mathbb{R}^n$, de forma que a matriz $P = \begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix}$ é invertível. Multiplicando ambos os lados de $AP = PD$, à direita, por P^{-1}, obtemos $A = PDP^{-1}$. □

Como já foi dito, os teoremas 8.25 e 8.25′ são, na verdade, formulações distintas do mesmo resultado. O teorema 8.25 tem a vantagem de ser mais sucinto. O teorema 8.25′, por sua vez, fornece mais informação. Ele não apenas caracteriza as matrizes diagonalizáveis, como também esboça uma "receita" para obter uma fatoração da forma $A = PDP^{-1}$: basta determinar uma base $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ de $\mathbb{R}^n$

composta por autovetores de A (se tal base existir) e, em seguida, "montar" as matrizes P e D de acordo com as equações (8.13) e (8.14).

A seguir, descrevemos esse processo de forma mais detalhada. Consideramos a diagonalização da matriz

$$A = \begin{bmatrix} 4 & 1 & -2 \\ 0 & 5 & 0 \\ -1 & 1 & 3 \end{bmatrix} \tag{8.19}$$

do exemplo 8.20. O roteiro dado abaixo, no entanto, é "geral", isto é, pode ser aplicado a uma matriz quadrada qualquer.

Passo 1: Determine os autovalores da matriz A.

O teorema 8.12 diz que os autovalores de A são as raízes de seu polinômio característico (reveja os exemplos 8.13 e 8.14). Não é fácil, em geral, achar as raízes de um polinômio de grau maior que 2, como mencionamos na seção 8.2. Nos exercícios deste texto, as matrizes de tamanho maior do que 2×2 serão "especiais" (triangulares, por exemplo) ou, então, seus autovalores serão dados.

No presente exemplo, o polinômio característico de A é dado por

$$p_A(x) = \det(A - xI) = \det \begin{bmatrix} 4-x & 1 & -2 \\ 0 & 5-x & 0 \\ -1 & 1 & 3-x \end{bmatrix} =$$

$$= (5-x) \cdot \big[(4-x)(3-x) - (-2)(-1)\big] = -(x-5)(x^2 - 7x + 10).$$

Utilizamos, acima, a expansão de Laplace do determinante[8] com respeito à segunda linha de $A - xI$. Verifique que $x^2 - 7x + 10 = (x-5)(x-2)$, de modo que o polinômio característico de A pode ser escrito na forma

$$p_A(x) = -(x-5)^2(x-2).$$

Os autovalores da matriz A são as raízes desse polinômio, ou seja, são:

$$\lambda_1 = 5 \qquad \text{e} \qquad \lambda_2 = 2.$$

Observe que $\lambda_1 = 5$ tem multiplicidade algébrica igual a 2.

Passo 2: Obtenha uma base para cada autoespaço de A.

A matriz A de (8.19) tem *dois* autoespaços, $E(A, \lambda_1)$ e $E(A, \lambda_2)$, associados, respectivamente, aos autovalores $\lambda_1 = 5$ e $\lambda_2 = 2$.

Já sabemos como obter uma base para cada subespaço: reveja os exemplos contidos na seção 8.3. No exemplo 8.20, em particular, obtivemos uma base do autoespaço $E(A, \lambda_1)$ dada pelo conjunto $\mathcal{B}_1 = \{\mathbf{v}_1, \mathbf{v}_2\}$, onde

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad \text{e} \quad \mathbf{v}_2 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}.$$

[8] Veja o exercício resolvido R6.6.

Observe, a propósito, que o autoespaço $E(A, \lambda_1)$ tem dimensão igual a 2.

Para determinar uma base de $E(A, \lambda_2) = \operatorname{Nuc}(A - \lambda_2 I) = \operatorname{Nuc}(A - 2I)$, resolvemos o sistema $(A - 2I)\mathbf{x} = \mathbf{0}$. A matriz completa desse sistema é

$$\begin{bmatrix} A - 2I \mid \mathbf{0} \end{bmatrix} = \left[\begin{array}{ccc|c} 4-2 & 1 & -2 & 0 \\ 0 & 5-2 & 0 & 0 \\ -1 & 1 & 3-2 & 0 \end{array}\right] = \left[\begin{array}{ccc|c} 2 & 1 & -2 & 0 \\ 0 & 3 & 0 & 0 \\ -1 & 1 & 1 & 0 \end{array}\right].$$

Verifique, via escalonamento, que o conjunto-solução do sistema $(A - 2I)\mathbf{x} = \mathbf{0}$ é descrito por

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_3 \\ 0 \\ x_3 \end{bmatrix} = x_3 \underbrace{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}}_{\mathbf{v}_3}.$$

Assim, obtemos uma base de $E(A, \lambda_2)$ dada por $\mathcal{B}_2 = \{\mathbf{v}_3\}$, onde $\mathbf{v}_3$ é o vetor indicado acima. Observe que $\dim E(A, \lambda_2) = 1$.

Passo 3: Determine se a matriz A é diagonalizável.

Os conjuntos $\mathcal{B}_1$ e $\mathcal{B}_2$ são bases dos autoespaços $E(A, \lambda_1)$ e $E(A, \lambda_2)$, respectivamente, com $\lambda_1 \neq \lambda_2$. Pelo corolário 8.23, o conjunto

$$\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$$

é linearmente independente. Este conjunto é, de fato, uma *base* de $\mathbb{R}^3$, visto que tem *três* elementos (veja a observação 4.27). Em suma, $\mathcal{B}$ é uma base de $\mathbb{R}^3$ composta por autovetores de A. Pelo teorema 8.25, a matriz A é diagonalizável.

O argumento acima pode parecer um pouco intrincado. É desejável estabelecer critérios mais objetivos para determinar se uma dada matriz é diagonalizável. Grosso modo, podemos afirmar que A é diagonalizável porque é uma matriz 3×3 e obtivemos *três* autovetores linearmente independentes, no passo 2 deste roteiro. Se A não fosse diagonalizável, teríamos obtido *menos do que três* vetores. Esse critério é comumente utilizado, na prática. Vamos reformulá-lo, de um modo mais preciso, no teorema 8.28, logo adiante.

Atenção: se a matriz dada não fosse diagonalizável, o processo terminaria aqui. O passo a seguir só faz sentido para matrizes diagonalizáveis!

Passo 4: Construa as matrizes P e D.

Já estabelecemos que a matriz A é diagonalizável. Resta, agora, escrever matrizes P e D na forma indicada pelas equações (8.13) e (8.14). O teorema 8.25′ garantirá, assim, a validade da relação $A = PDP^{-1}$.

Primeiro, escrevemos a matriz P, conforme a equação (8.13), usando os elementos da base $\mathcal{B}$, ou seja, usando os autovetores $\mathbf{v}_1$, $\mathbf{v}_2$ e $\mathbf{v}_3$ obtidos no passo 2:

$$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} = \begin{bmatrix} 1 & -2 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

A ordem em que escrevemos os autovetores é, por enquanto, irrelevante. Note que P é uma matriz invertível, pois P é quadrada e suas colunas são linearmente independentes, como observamos no passo 3.

Em seguida, escrevemos a matriz diagonal D, conforme a equação (8.14). É fundamental, agora, que a ordem dos autovalores na diagonal de D seja condizente com a ordem adotada para as colunas de P, como estabelece o teorema 8.25′. Lembre, do passo 2, que as *duas primeiras* colunas de P ($\mathbf{v}_1$ e $\mathbf{v}_2$) são autovetores associados ao autovalor $\lambda_1 = 5$. A *terceira* coluna de P ($\mathbf{v}_3$) é um autovetor associado a $\lambda_2 = 2$. Assim, temos

$$D = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Convém salientar que o autovalor $\lambda_1 = 5$ aparece *duas vezes* na diagonal de D. Isso ocorre porque a dimensão do autoespaço $E(A, \lambda_1)$ é igual a 2, como indicamos no passo 2.

Isso termina o processo de diagonalização da matriz A. É recomendável conferir o trabalho realizado: calcule a inversa P^{-1} da matriz P e verifique, diretamente, que vale a relação $A = PDP^{-1}$. Veja também o exercício **??**.

Observação 8.26
O exemplo acima mostra que os autovalores $\lambda_1, \ldots, \lambda_n$ indicados no enunciado do teorema 8.25′, que aparecem na diagonal da matriz D, não são necessariamente distintos. Às vezes, é conveniente escrever os autovalores de uma matriz *repetidos de acordo com as suas multiplicidades*. Os autovalores da matriz A considerada acima seriam escritos, dessa maneira, como $\lambda_1 = 5$, $\lambda_1 = 5$, $\lambda_2 = 2$.

Observação 8.27
A diagonalização de uma matriz, quando existe, *não é única*. Em outras palavras, se uma matriz A é diagonalizável, então existe mais de uma maneira de fatorá-la na forma $A = PDP^{-1}$. Veja os exercícios ??? e ???.

No passo 3 do roteiro acima, esboçamos um "critério prático" para determinar se uma dada matriz é diagonalizável. O teorema a seguir formula esse critério de uma maneira mais rigorosa. Esse resultado está contido no teorema 8.32, que iremos enunciar e demonstrar mais adiante.

Teorema 8.28
Uma matriz $n \times n$ será diagonalizável se, e somente se, a soma das dimensões de seus autoespaços for igual a n.

É fácil aplicar o teorema 8.28 ao passo 3 do exemplo acima: os autoespaços $E(A, \lambda_1)$ e $E(A, \lambda_2)$ da matriz A têm dimensões 2 e 1, respectivamente, como observamos no passo 2. A soma das dimensões, assim, é 3. Como A é uma matriz 3×3, o teorema 8.28 garante que A é diagonalizável.

Vejamos mais um exemplo de diagonalização de uma matriz.

Exemplo 8.29

Diagonalize a matriz $B = \begin{bmatrix} 17 & -30 \\ 9 & -16 \end{bmatrix}$, se possível.

Solução: Vamos seguir o roteiro dado acima.

Passo 1: achar os autovalores de B.

O polinômio característico de B é

$$p_B(x) = \det(B - xI) = \det \begin{bmatrix} 17 - x & -30 \\ 9 & -16 - x \end{bmatrix} =$$

$$= (17 - x) \cdot (-16 - x) - (-30) \cdot 9 = x^2 - x - 2 = (x - 2)(x + 1).$$

Os autovalores de B são as raízes de $p_B(x)$, ou seja, são:

$$\lambda_1 = 2 \qquad \text{e} \qquad \lambda_2 = -1.$$

Como $p_B(x) = x^2 - x - 2$ é um polinômio de grau 2, poderíamos, também, ter aplicado a "fórmula de Báskara" para obter estas raízes.

Passo 2: obter uma base para cada autoespaço de B.

Para achar uma base de $E(B, \lambda_1) = \mathrm{Nuc}(B - \lambda_1 I) = \mathrm{Nuc}(B - 2I)$, resolvemos o sistema $(B - 2I)\mathbf{x} = \mathbf{0}$. Complete os passos do escalonamento da matriz completa desse sistema:

$$\begin{bmatrix} B - 2I \mid \mathbf{0} \end{bmatrix} = \left[\begin{array}{cc|c} 17 - 2 & -30 & 0 \\ 9 & -16 - 2 & 0 \end{array}\right] = \left[\begin{array}{cc|c} 15 & -30 & 0 \\ 9 & -18 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} 1 & -2 & 0 \\ 0 & 0 & 0 \end{array}\right].$$

Dessa maneira, o conjunto-solução de $(B - 2I)\mathbf{x} = \mathbf{0}$ é descrito por

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} 2 \\ 1 \end{bmatrix},$$

e, portanto, $\{\mathbf{v}_1\} = \left\{ \left[\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right] \right\}$ é uma base de $E(B, \lambda_1)$.

Para obter uma base de $E(B, \lambda_2) = \mathrm{Nuc}(B - \lambda_2 I) = \mathrm{Nuc}(B + I)$, resolvemos o sistema $(B + I)\mathbf{x} = \mathbf{0}$. Verifique, via escalonamento, que seu conjunto-solução é descrito por

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{5}{3}x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} 5/3 \\ 1 \end{bmatrix}.$$

Considerando $x_2 = 3$ (para evitar trabalhar com frações), obtemos uma base de $E(B, \lambda_2)$ dada por $\{\mathbf{v}_2\} = \left\{ \left[\begin{smallmatrix} 5 \\ 3 \end{smallmatrix}\right] \right\}$.

Passo 3: determinar se a matriz B é diagonalizável.

Pelo teorema 8.22, o conjunto $\{\mathbf{v}_1, \mathbf{v}_2\}$ é linearmente independente, visto que $\mathbf{v}_1$ e $\mathbf{v}_2$ são autovetores de B associados a *autovalores distintos*. Conforme a observação 4.27, este conjunto é, de fato, uma *base* de $\mathbb{R}^2$, já que contém *dois* vetores. Aplicando o teorema 8.25, concluímos que a matriz B é diagonalizável.

Alternativamente, podemos aplicar o "critério prático" do teorema 8.28: os autoespaços $E(B, \lambda_1)$ e $E(B, \lambda_2)$ têm, cada um, dimensão igual a 1 (explique!).

A soma das dimensões, portanto, é igual a 2. Visto que B é 2×2, o teorema 8.28 afirma que B é diagonalizável. Simples, não?

Passo 4: escrever matrizes P e D tais que $B = PDP^{-1}$.

Mais uma vez, basta usar as equações (8.13) e (8.14). Escrevemos, primeiro, a matriz P, usando os autovetores $\mathbf{v}_1$ e $\mathbf{v}_2$ obtidos no passo 2:

$$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}.$$

Em seguida, escrevemos a matriz diagonal D, usando os autovalores correspondentes às colunas de P:

$$D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix}.$$

Isso termina o processo de diagonalização: o teorema 8.25′ garante que vale a relação de semelhança $B = PDP^{-1}$. Verifique! □

Vejamos, agora, exemplos de matrizes que *não* são diagonalizáveis.

Exemplo 8.30

Diagonalize a matriz $G = \begin{bmatrix} 4 & 1 \\ -1 & 2 \end{bmatrix}$, se possível.

Solução: O polinômio característico de G é

$$p_G(x) = \det(G - xI) = \det \begin{bmatrix} 4 - x & 1 \\ -1 & 2 - x \end{bmatrix} =$$

$$= (4 - x) \cdot (2 - x) - 1 \cdot (-1) = x^2 - 6x + 9 = (x - 3)^2.$$

Assim, a matriz G possui unicamente o autovalor $\lambda = 3$, com multiplicidade 2. Consequentemente, G possui apenas o autoespaço $E(G, \lambda)$. Para encontrar uma base desse autoespaço, vamos resolver o sistema $(G - 3I)\mathbf{x} = \mathbf{0}$:

$$\begin{bmatrix} G - 3I \mid \mathbf{0} \end{bmatrix} = \left[\begin{array}{cc|c} 4-3 & 1 & 0 \\ -1 & 2-3 & 0 \end{array}\right] = \left[\begin{array}{cc|c} 1 & 1 & 0 \\ -1 & -1 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right].$$

O conjunto-solução do sistema $(G - 3I)\mathbf{x} = \mathbf{0}$ é descrito por

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix},$$

e, assim, $\{\mathbf{v}\} = \left\{ \left[\begin{smallmatrix} -1 \\ 1 \end{smallmatrix} \right] \right\}$ é uma base de $E(G, \lambda)$. Isso conclui os passos 1 e 2 do processo de diagonalização.

Perceba que o autoespaço $E(G, \lambda)$ tem dimensão igual a 1. Este é o único autoespaço de G, como mencionamos acima. A "soma" das dimensões dos autoespaços, dessa forma, é igual a 1. Por outro lado, G é uma matriz 2×2. Aplicando o critério do teorema 8.28, concluímos que *G não é diagonalizável.*

É possível chegar à mesma conclusão argumentando, diretamente, que não existe uma base de $\mathbb{R}^2$ composta por autovetores de G, e aplicando o teorema 8.25. Veja o exercício **??**. □

Exemplo 8.31

Diagonalize a matriz $C = \begin{bmatrix} 3 & -2 \\ 2 & 3 \end{bmatrix}$ do exemplo 8.16, se possível.

Solução: Como vimos no exemplo 8.16, as raízes do polinômio característico de C são os números complexos conjugados $\lambda_1 = 3 + 2i$ e $\lambda_2 = 3 - 2i$. A matriz C não possui autovalores reais e, portanto, não possui autovetor algum em $\mathbb{R}^2$. Em decorrência do teorema 8.25, podemos concluir imediatamente que *a matriz C não é diagonalizável*. □

O teorema a seguir contém o teorema 8.28, e aprofunda o nosso conhecimento sobre a estrutura de operadores diagonalizáveis.

DIZER QUE, NO CONTEXTO DESSA SEÇÃO, DIAGONALIZAR SIGNIFICA "DIAGONALIZAR SOBRE OS REAIS". EXPLICAR O QUE SIGNIFICA ISSO.

MELHORAR UM POUCO O ENUNCIADO ABAIXO??? E PROVÁ-LO!!!

Teorema 8.32

Seja A uma matriz $n \times n$. As seguintes afirmativas são equivalentes:

(a) *A é diagonalizável (sobre os reais).*

(b) *A soma das dimensões dos autoespaços (reais) de A é n.*

(c) *O polinômio característico de A tem apenas raízes reais e a multiplicidade geométrica de cada autovalor de A é igual à sua multiplicidade algébrica.*

A FAZER:

- Demonstração do teorema (no final do capítulo?).

- Exemplo: A matriz B do exemplo 8.14 não é diagonalizável: a dimensão de $E(A, \lambda_2)$ é 1; a soma das dimensões dos autoespaços é 3.

Recorde que, para cada autovalor λ de uma matriz A, vale $1 \leqslant \dim E(A, \lambda) \leqslant n_\lambda$, onde n_λ representa a multiplicidade algébrica de λ. Desta observação e do teorema acima, segue o seguinte resultado.

Corolário 8.33

Se uma matriz $n \times n$ tiver n autovalores reais e distintos, então ela será diagonalizável.

A FAZER:

- Dar um exemplo 3×3???

- OBS: Cuidado: o corolário acima fornece uma condição *suficiente*, porém não *necessária* para a "diagonalizabilidade" de uma matriz. Considere, por exemplo, a matriz A dada em (8.19). Observe que A é uma matriz 3×3 com apenas 2 autovalores distintos ($\lambda_1 = 5$ e $\lambda_2 = 2$). No entanto, a matriz A é diagonalizável, como vimos.

8.5 Matrizes reais com autovalores complexos

BLAH BLAH ???

8.6 Classificação de matrizes 2×2

BLAH BLAH ???

8.7 Demonstrações dos resultados sobre autoespaços (leitura opcional)

Fornecemos, abaixo, as demonstrações dos resultados da seção 8.3.

Demonstração do teorema 8.21: Considere qualquer autovalor λ da matriz A. Vamos denotar sua multiplicidade algébrica por n_λ e sua multiplicidade geométrica por p. Lembre que, de acordo com as respectivas definições, n_λ é a multiplicidade de λ enquanto raiz do polinômio característico de A, e $p = \dim E(A, \lambda)$. Queremos provar que $p \leqslant n_\lambda$.

Seja $\mathcal{I} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ uma base de $E(A, \lambda)$. Observe que o número de elementos de $\mathcal{I}$ corresponde à dimensão p de $E(A, \lambda)$, conforme a definição 4.24. Esse fato é crucial na demonstração. Observe, também, que $\mathcal{I}$ é um conjunto linearmente independente de vetores de $\mathbb{R}^n$. Pelo teorema 4.19, é possível obter uma base $\mathbb{R}^n$ por acrescentar a $\mathcal{I}$ certos vetores $\mathbf{v}_{p+1}$, $\mathbf{v}_{p+2}$, $\ldots$, $\mathbf{v}_n$. Assim, construímos uma base $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ de $\mathbb{R}^n$ cujos primeiros p elementos são autovetores de A associados ao autovalor λ.

Vamos, agora, usar ideias da seção 7.2. Seja $T : \mathbb{R}^n \to \mathbb{R}^n$ o operador linear dado por $\mathbf{x} \mapsto A\mathbf{x}$, e considere a matriz $[T]_\mathcal{B}$ desse operador com respeito à base $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. Lembre que $[T]_\mathcal{B}$ é a matriz $n \times n$ cuja j-ésima coluna é dada pelo vetor de coordenadas de $T(\mathbf{v}_j)$ com respeito à base $\mathcal{B}$ (veja as equações (**??**) e (**??**)). Visto que $\mathbf{v}_1$, $\ldots$, $\mathbf{v}_p$ são autovetores de A associados a λ, valem

$$T(\mathbf{v}_1) = A\mathbf{v}_1 = \lambda\mathbf{v}_1, \quad T(\mathbf{v}_2) = A\mathbf{v}_2 = \lambda\mathbf{v}_2, \quad \ldots\ , \quad T(\mathbf{v}_p) = A\mathbf{v}_p = \lambda\mathbf{v}_p.$$

As equações acima mostram que, para $j = 1, \ldots, p$, a j-ésima coordenada do vetor $T(\mathbf{v}_j)$ com respeito à base $\mathcal{B}$ é λ, e as demais coordenadas são todas iguais a zero.[9] Assim, a matriz do operador T com respeito à base $\mathcal{B}$ é da forma

$$[T]_\mathcal{B} = \left[\begin{array}{cccc|cccc} \lambda & 0 & \cdots & 0 & a_{1,p+1} & a_{1,p+2} & \cdots & a_{1,n} \\ 0 & \lambda & \cdots & 0 & a_{2,p+1} & a_{2,p+2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda & a_{p,p+1} & a_{p,p+2} & \cdots & a_{p,n} \\ \hline 0 & 0 & \cdots & 0 & a_{p+1,p+1} & a_{p+1,p+2} & \cdots & a_{p+1,n} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & a_{n,p+1} & a_{n,p+2} & \cdots & a_{n,n} \end{array}\right] \begin{array}{l} \left.\vphantom{\begin{array}{c}a\\a\\ \vdots\\a\end{array}}\right\} p \text{ linhas} \\ \left.\vphantom{\begin{array}{c}a\\ \vdots\\a\end{array}}\right\} n-p \text{ linhas.} \end{array} \tag{8.20}$$

$$\underbrace{\qquad\qquad}_{p \text{ colunas}} \qquad \underbrace{\qquad\qquad\qquad\qquad}_{n-p \text{ colunas}}$$

[9] Recorde a definição 4.32: as coordenadas de um vetor $\mathbf{u}$ de $\mathbb{R}^n$ com respeito à base $\mathcal{B}$ são os pesos na expressão de $\mathbf{u}$ como combinação linear dos vetores $\mathbf{v}_1$, $\ldots$, $\mathbf{v}_n$ de $\mathcal{B}$.

Perceba, a propósito, que nada afirmamos sobre as últimas $n-p$ colunas da matriz $[T]_{\mathcal{B}}$. Meramente indicamos as suas entradas por $a_{i,j}$, na equação acima.[10] O valor dessas entradas, para $j > p$, é irrelevante na demonstração.

Podemos reescrever a equação (8.20), simplificadamente, como

$$[T]_{\mathcal{B}} = \left[\begin{array}{c|c} \Lambda & G \\ \hline \mathit{0} & H \end{array}\right], \tag{8.21}$$

onde Λ é a matriz diagonal $p \times p$ contendo o autovalor λ em cada entrada da diagonal, $\mathit{0}$ é a matriz zero de tamanho $n-p \times p$, e G e H são submatrizes de tamanho $p \times n-p$ e $n-p \times n-p$, respectivamente. Como mencionamos acima, o valor das entradas de G e de H é irrelevante na demonstração.

A seguir, vamos calcular o polinômio característico da matriz $[T]_{\mathcal{B}}$. Conforme a observação ???, as matrizes A e $[T]_{\mathcal{B}}$ são semelhantes, visto que são associadas ao mesmo operador (lembre que A é a matriz canônica do operador T). Pelo teorema 8.17, A e $[T]_{\mathcal{B}}$ têm o mesmo polinômio característico.

Usando a notação da equação (8.21), temos

$$[T]_{\mathcal{B}} - xI_n = \left[\begin{array}{c|c} \Lambda & G \\ \hline \mathit{0} & H \end{array}\right] - x \left[\begin{array}{c|c} I_p & \mathit{0} \\ \hline \mathit{0} & I_{n-p} \end{array}\right] = \left[\begin{array}{c|c} \Lambda - xI_p & G \\ \hline \mathit{0} & H - xI_{n-p} \end{array}\right],$$

onde I_k representa a matriz identidade $k \times k$, para $k \in \mathbb{N}$. Observe que $[T]_{\mathcal{B}} - xI_n$ é uma *matriz triangular superior em blocos*. Pelo teorema **??**,

$$\det\left([T]_{\mathcal{B}} - xI_n\right) = \det\left(\Lambda - xI_p\right) \cdot \det\left(H - xI_{n-p}\right).$$

Em outras palavras, o polinômio característico de $[T]_{\mathcal{B}}$ é dado pelo produto dos polinômios característicos de Λ e de H:

$$p_{[T]_{\mathcal{B}}}(x) = p_{\Lambda}(x) \cdot p_H(x).$$

É fácil ver que $p_{\Lambda}(x) = (\lambda - x)^p$. Deixamos a verificação desse fato como exercício (note que $\Lambda = \lambda I_p$). Lembrando que os polinômios característicos de A e de $[T]_{\mathcal{B}}$ coincidem, concluimos que

$$p_A(x) = p_{[T]_{\mathcal{B}}}(x) = (\lambda - x)^p \cdot p_H(x).$$

A equação acima mostra que a multiplicidade de λ enquanto raiz do polinômio $p_A(x)$ é, *no mínimo*, igual a p. (Veja o exercício ???.) Em outras palavras, a multiplicidade n_λ é maior que ou igual a p, como queríamos provar. $\square$

Demonstração do teorema 8.22: Queremos mostrar que $\mathcal{G} = \{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ é um conjunto linearmente independente. É útil lembrar que $\mathbf{v}_1, \dots, \mathbf{v}_m$ são todos *não-nulos*, já que são autovetores de A. Além disso, podemos admitir que $m \geqslant 2$, já que o resultado é evidente no caso $m = 1$.

[10] Como mencionamos, e conforme a teoria da seção ??, a entrada $a_{i,j}$ é dada pela i-ésima coordenada do vetor $T(\mathbf{v}_j)$ com respeito à base $\mathcal{B}$.

A estratégia da demonstração será a seguinte: vamos supor, contrariamente ao que queremos provar, que $\mathcal{G}$ seja um conjunto linearmente *dependente*. Sob as hipóteses do teorema, essa suposição levará a uma contradição, isto é, a uma impossibilidade lógica. Isso provará o resultado desejado.[11]

Seja $V = \operatorname{Span} \mathcal{G}$. Pelo teorema 4.18, existe uma base $\mathcal{B}$ de V dada por um subconjunto de $\mathcal{G}$. Mediante um possível rearranjo dos índices em $\mathcal{G}$, podemos admitir, sem perda de generalidade, que $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$.[12] Em princípio, o caso $\mathcal{B} = \mathcal{G}$ seria possível, de acordo com o teorema 4.18. No entanto, se o conjunto $\mathcal{G}$ é linearmente dependente, como supusemos, então $\mathcal{G}$ não pode ser, ele mesmo, uma base de V. Assim, a base $\mathcal{B}$ é, de fato, um *subconjunto próprio* de $\mathcal{G}$, ou seja, $\mathcal{B}$ está *estritamente* contido em $\mathcal{G}$. Em particular, o número p de elementos de $\mathcal{B}$ é menor do que m. Observe, a propósito, que p representa a dimensão de V.

Como o vetor $\mathbf{v}_m$ pertence a V (veja a observação 2.15), podemos escrevê-lo como combinação linear dos vetores da base $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$:

$$\mathbf{v}_m = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p. \tag{8.22}$$

Multiplicando ambos os lados desta equação pela matriz A, obtemos

$$A\mathbf{v}_m = c_1A\mathbf{v}_1 + c_2A\mathbf{v}_2 + \cdots + c_pA\mathbf{v}_p.$$

Por hipótese, $\mathbf{v}_1, \ldots, \mathbf{v}_m$ são autovetores de A associados, respectivamente, a $\lambda_1, \ldots, \lambda_m$, de modo que $A\mathbf{v}_j = \lambda_j\mathbf{v}_j$ para cada j. Assim, a equação acima se reescreve como

$$\lambda_m\mathbf{v}_m = c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 + \cdots + c_p\lambda_p\mathbf{v}_p. \tag{8.23}$$

Multiplicando, agora, ambos os lados da equação (8.22) pelo escalar λ_m, obtemos

$$\lambda_m\mathbf{v}_m = c_1\lambda_m\mathbf{v}_1 + c_2\lambda_m\mathbf{v}_2 + \cdots + c_p\lambda_m\mathbf{v}_p. \tag{8.24}$$

Finalmente, subtraindo a equação (8.24) da (8.23), obtemos

$$\mathbf{0} = c_1(\lambda_1 - \lambda_m)\mathbf{v}_1 + c_2(\lambda_2 - \lambda_m)\mathbf{v}_2 + \cdots + c_p(\lambda_p - \lambda_m)\mathbf{v}_p. \tag{8.25}$$

Os vetores $\mathbf{v}_1, \ldots, \mathbf{v}_p$ são linearmente independentes, pois compõem uma base de V. Assim, todos os pesos $c_j(\lambda_j - \lambda_m)$ na equação (8.25) são necessariamente iguais a zero. Por hipótese, os autovalores λ_j são distintos, de modo que nenhum dos fatores $\lambda_j - \lambda_m$ pode ser igual a zero. Isso mostra que os escalares c_j têm que ser todos nulos. Sendo assim, a equação (8.22) diz que $\mathbf{v}_m$ é o vetor zero. Isso, contudo, contradiz a hipótese de que $\mathbf{v}_m$ é um autovetor de A.

Recapitulando: a suposição de que o conjunto $\mathcal{G} = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ é linearmente *dependente* é logicamente incompatível com as hipóteses do teorema. Isso prova, então, que o conjunto $\mathcal{G}$ tem que ser *independente*, como queríamos provar. $\square$

[11] Este tipo de argumento chama-se "prova por contradição" ou "redução ao absurdo" e pode parecer, à primeira vista, um pouco anti-intuitivo.

[12] Em outras palavras, podemos reordenar os vetores do conjunto $\mathcal{G} = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$, se necessário, de maneira que os *primeiros* p vetores componham uma base $\mathcal{B}$ para o subespaço V gerado por todos os m vetores.

Demonstração do corolário 8.23: Vamos usar a notação introduzida logo após o enunciado do corolário, isto é, vamos denotar os vetores da base $\mathcal{B}_k$ de $E(A, \lambda_k)$ por $\mathbf{v}_1^k, \ldots, \mathbf{v}_{p_k}^k$, de forma que

$$\mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_m = \{\mathbf{v}_1^1, \mathbf{v}_2^1, \ldots, \mathbf{v}_{p_1}^1, \mathbf{v}_1^2, \mathbf{v}_2^2, \ldots, \mathbf{v}_{p_2}^2, \ldots, \mathbf{v}_1^m, \mathbf{v}_2^m, \ldots, \mathbf{v}_{p_m}^m\},$$

conforme a equação (8.12). Considere a relação

$$\overbrace{x_1^1\mathbf{v}_1^1 + x_2^1\mathbf{v}_2^1 + \cdots + x_{p_1}^1\mathbf{v}_{p_1}^1}^{\mathbf{u}^1} + \overbrace{x_1^2\mathbf{v}_1^2 + x_2^2\mathbf{v}_2^2 + \cdots + x_{p_2}^2\mathbf{v}_{p_2}^2}^{\mathbf{u}^2} + \cdots +$$
$$+ \underbrace{x_1^m\mathbf{v}_1^m + x_2^m\mathbf{v}_2^m + \cdots + x_{p_m}^m\mathbf{v}_{p_m}^m}_{\mathbf{u}^m} = \mathbf{0}. \quad (8.26)$$

Para mostrar que os vetores $\mathbf{v}_j^k$ são linearmente independentes, vamos mostrar que todos os pesos x_j^k na relação acima são, necessariamente, iguais a zero.[13]

Vamos denotar o vetor $x_1^k\mathbf{v}_1^k + \cdots + x_{p_k}^k\mathbf{v}_{p_k}^k$ por $\mathbf{u}^k$, para $k = 1, \ldots, m$, como indicado em (8.26). Assim, podemos reescrever essa equação como

$$\mathbf{u}^1 + \mathbf{u}^2 + \cdots + \mathbf{u}^m = \mathbf{0}. \quad (8.27)$$

Observe que cada $\mathbf{u}^k$ pertence ao autoespaço $E(A, \lambda_k)$, pois é uma combinação linear de vetores da base $\mathcal{B}_k$. Assim, ou $\mathbf{u}^k = \mathbf{0}$, ou, então, $\mathbf{u}^k$ é um autovetor associado a λ_k. Contudo, se algum dos vetores $\mathbf{u}^k$ fosse não-nulo, então (8.27) representaria uma relação de dependência linear entre autovetores associados a autovalores distintos, contrariando o teorema 8.22.

Concluímos, assim, que $\mathbf{u}^k = x_1^k\mathbf{v}_1^k + \cdots + x_{p_k}^k\mathbf{v}_{p_k}^k = \mathbf{0}$, para $k = 1, \ldots, m$. Os vetores $\mathbf{v}_1^k, \ldots, \mathbf{v}_{p_k}^k$ são linearmente independentes, já que compõem a base $\mathcal{B}_k$. Assim, os escalares $x_1^k, \ldots, x_{p_k}^k$ são necessariamente iguais a zero. Isso termina a prova de que todos os pesos na relação (8.26) são iguais a zero. $\square$

Exercícios resolvidos

R8.1. MOSTRAR POR INDUCAO QUE VALE A FÓRMULA (8.1).

R8.2. Sejam A, B e P matrizes $n \times n$, e suponha que P seja invertível. Mostre que a relação $A = PBP^{-1}$ é equivalente a $P^{-1}AP = B$.

Solução: Supondo que $A = PBP^{-1}$, temos

$$A = PBP^{-1} \implies P^{-1}AP = P^{-1}(PBP^{-1})P$$
$$\implies P^{-1}AP = IBI \implies P^{-1}AP = B.$$

Reciprocamente, supondo que $P^{-1}AP = B$, temos

$$P^{-1}AP = B \implies P(P^{-1}AP)P^{-1} = PBP^{-1}$$
$$\implies IAI = PBP^{-1} \implies A = PBP^{-1}. \quad \square$$

[13] Recorde a definição 3.14 e o teorema 3.21. Veja, em particular, a afirmativa (d) do teorema.

Observação: Este exercício está ligado ao conceito de *semelhança de matrizes*. Recorde a definição 7.11.

Exercícios propostos

P8.1. Seja A uma matriz de tamanho $2{\times}3$. A matriz A^2 está definida? Justifique.

P8.2. Seja $A = [a_{ij}]$ a matriz do exemplo 8.2. Verifique que A^2 *não coincide* com a matriz $[a_{ij}^2]$ obtida tomando o quadrado de cada entrada de A.

P8.3. Seja A, novamente, a matriz do exemplo 8.2. Calcule A^k diretamente, multiplicando A por ela própria repetidas vezes, para alguns valores de k. Considere, por exemplo, os casos $k = 1$, 2, 3 e 4. Em seguida, aplique a fórmula (8.4) em cada caso, e verifique que os resultados coincidem.

P8.4. (a) Seja A a matriz do exemplo 8.2. Verifique que a equação $A\mathbf{v} = \lambda\mathbf{v}$ equivale ao sistema

$$\begin{cases} 8v_1 + 5v_2 - \lambda v_1 = 0 \\ -10v_1 - 7v_2 - \lambda v_2 = 0, \end{cases}$$

nas variáveis λ, v_1 e v_2 (onde v_1 e v_2 são as coordenadas de $\mathbf{v}$). Note que esse sistema é *não-linear*, em vista dos produtos λv_1 e λv_2.

(b) Observe que, em geral, se A é uma matriz $n \times n$, a equação $A\mathbf{x} = \lambda\mathbf{x}$ representa um sistema não-linear nas variáveis λ, $x_1, \ldots, x_n$, onde $x_1, \ldots, x_n$ são as coordenadas do vetor $\mathbf{x}$.

Capítulo 9

Produto Interno, Projeções e Operadores Ortogonais

9.1 Introdução

Até o momento tratamos de diversos conceitos do $\mathbb{R}^n$, sem abordar os conceitos de norma de vetor e de ângulo entre vetores e suas consequências diretas, as quais são a distância entre pontos e a ideia de perpendicularidade. Para abordar esses conceitos, precisamos definir o conceito de produto escalar. Aproveitando o momento, vamos introduzir um conceito mais geral, que é o de produto interno. O produto interno nos permitirá estender as noções de distância e perpendicularidade a outros espaços vetoriais diferentes do $\mathbb{R}^n$.

Introduzindo o conceito de produto interno, podemos mostrar que o produto escalar do $\mathbb{R}^n$ é um exemplo de produto interno. Por simplicidade, é bom imaginar que, pelo menos no princípio, quando falamos de produto interno, nos referimos ao produto escalar do $\mathbb{R}^n$.

9.2 Produto Interno

Iniciemos com a definição do produto interno no $\mathbb{R}^n$.

Definição 9.1
Suponha que, para cada par de vetores $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, existe uma função que associa um número real, denotado por $\langle \mathbf{u}, \mathbf{v} \rangle$. Essa função é chamada **produto interno** de V se satisfizer às seguintes propriedades:

I_1 (Linearidade) $\langle \alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{w} \rangle = \alpha \langle \mathbf{u}, \mathbf{w} \rangle + \beta \langle \mathbf{v}, \mathbf{w} \rangle$;

I_2 (Simetria) $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$;

I_3 (Positividade) $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$, e $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ se, e somente se, $\mathbf{u} = 0$, para todo $\mathbf{u}, \mathbf{v}$ e $\mathbf{w}$ em $\mathbb{R}^n$ e $\alpha, \beta \in \mathbb{R}$. O $\mathbb{R}^n$ com o produto interno é denominado **espaço vetorial com produto interno** ou ainda **espaço euclidiano**[1].

Vamos definir o exemplo mais importante de produto interno no $\mathbb{R}^n$.

Exemplo 9.2
O produto escalar usual do $\mathbb{R}^2$, isto é, para $\mathbf{u} = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]$ e $\mathbf{v} = \left[\begin{smallmatrix} y_1 \\ y_2 \end{smallmatrix}\right]$ definido por

$$\langle \mathbf{u}, \mathbf{v} \rangle = x_1 y_1 + x_2 y_2,$$

é um exemplo de produto interno. De fato, é claro que esta é uma função que toma dois vetores e retorna um número. Vamos verificar que esta função satisfaz às propriedades para que seja um produto interno. Para verificar a linearidade, considere ainda o vetor $\mathbf{w} = \left[\begin{smallmatrix} z_1 \\ z_2 \end{smallmatrix}\right]$. Então,

$$\begin{aligned}
\langle \alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{w} \rangle &= \left\langle \alpha \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \beta \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \right\rangle \\
&= \left\langle \begin{bmatrix} \alpha x_1 + \beta y_1 \\ \alpha x_2 + \beta y_2 \end{bmatrix}, \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \right\rangle = (\alpha x_1 + \beta y_1) z_1 + (\alpha x_2 + \beta y_2) z_2 \\
&= \alpha (x_1 z_1 + x_2 z_2) + \beta (y_1 z_1 + y_2 z_2) = \alpha \langle \mathbf{u}, \mathbf{w} \rangle + \beta \langle \mathbf{v}, \mathbf{w} \rangle .
\end{aligned}$$

A simetria é evidente e, para a positividade, basta observar que $\langle \mathbf{u}, \mathbf{u} \rangle = x_1^2 + x_2^2 \geq 0$ é zero, se e somente se, $x_1 = x_2 = 0$, isto é, no caso em que $\mathbf{u} = \mathbf{0}$.

Podemos generalizar o produto escalar usual para o $\mathbb{R}^n$, com $n > 2$.

Exemplo 9.3
Considere os vetores $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. Digamos que

$$\mathbf{u} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \mathbf{v} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \text{defina } \langle \mathbf{u}, \mathbf{v} \rangle = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

Verifique que a função assim definida satisfaz às propriedades exigidas de um produto interno, como fizemos para o caso do produto escalar no plano.

Gostaria de frisar que, para o melhor entendimento, no primeiro momento é interessante quando ler que $\mathbb{R}^n$ é um espaço vetorial munido de um produto interno, imaginar que o produto interno é o produto escalar.

Vamos deduzir algumas propriedades que devem valer para qualquer produto interno. Para fazer isto só usaremos propriedades que constam na definição. Em particular, devem valer para o produto escalar.

[1] O produto interno pode ser generalizado para vetores de $\mathbb{C}^n$, mas para fazer isso, precisamos substituir a propriedade I_2 por $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$, onde a barra denota a conjugação complexa. Nesse texto faremos apenas a teoria para espaços vetoriais reais.

1. O produto interno também é linear na sua segunda entrada. Veja:

$$\begin{aligned}\langle \mathbf{u}, \alpha\mathbf{v} + \beta\mathbf{w}\rangle &= \langle \alpha\mathbf{v} + \beta\mathbf{w}, \mathbf{u}\rangle \\ &= \alpha\langle \mathbf{v}, \mathbf{u}\rangle + \beta\langle \mathbf{w}, \mathbf{u}\rangle = \alpha\langle \mathbf{u}, \mathbf{v}\rangle + \beta\langle \mathbf{u}, \mathbf{w}\rangle.\end{aligned}$$

A primeira e terceira igualdades seguem pela simetria e a segunda igualdade da linearidade.

2. O produto $\langle \mathbf{0}, \mathbf{u}\rangle = \langle 0\mathbf{0}, \mathbf{u}\rangle = 0\langle \mathbf{0}, \mathbf{u}\rangle = 0$, e, pela observação 1, temos, também, $\langle \mathbf{u}, \mathbf{0}\rangle = 0$.

Vamos dar um exemplo de uma função diferente do produto escalar que também satisfaz às propriedades do produto interno. E, mais adiante, vamos usar o produto interno para introduzir as noções de distância e ângulo entre vetores. Como podemos usar várias funções, vemos que existem várias maneiras (diferentes) de medir distância e ângulo entre vetores. Então, podemos fazer diversas perguntas, tais como: o que é invariante entre uma forma de medir e outra? E, sabendo as medidas de um objeto com respeito a um produto interno, como podemos determinar as suas medidas em relação a um outro produto interno? Com respeito à primeira pergunta, existe um ramo na matemática chamado topologia que responde a boa parte desta questão.

Exemplo 9.4
Sejam $\mathbf{u} = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]$ e $\mathbf{v} = \left[\begin{smallmatrix} y_1 \\ y_2 \end{smallmatrix}\right]$ em $\mathbb{R}^2$, e considere

$$\langle \mathbf{u}, \mathbf{v}\rangle = 2x_1y_1 - x_1y_2 - x_2y_1 + x_2y_2.$$

Vamos verificar que essa função é um produto interno no plano. Para isso, seja $\mathbf{w} = \left[\begin{smallmatrix} z_1 \\ z_2 \end{smallmatrix}\right]$.

Vamos verificar a bilinearidade. Para isso, considere $\mathbf{w} = \left[\begin{smallmatrix} z_1 \\ z_2 \end{smallmatrix}\right]$ e $\alpha \in \mathbb{R}$ um escalar. Então,

$$\begin{aligned}\langle \mathbf{u} + \alpha\mathbf{v}, \mathbf{w}\rangle &= \left\langle \left[\begin{smallmatrix} x_1+\alpha y_1 \\ x_2+\alpha y_2 \end{smallmatrix}\right], \left[\begin{smallmatrix} z_1 \\ z_2 \end{smallmatrix}\right]\right\rangle \\ &= 2(x_1 + \alpha y_1)z_1 - (x_1 + \alpha y_1)z_2 - (x_2 + \alpha y_2)z_1 + (x_2 + \alpha y_2)z_2 \\ &= 2x_1z_1 - x_1z_2 - x_2z_1 + x_2z_2 + \alpha(2y_1z_1 - y_1z_2 - y_2z_1 + y_2z_2) \\ &= \left\langle \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right], \left[\begin{smallmatrix} z_1 \\ z_2 \end{smallmatrix}\right]\right\rangle + \alpha\left\langle \left[\begin{smallmatrix} y_1 \\ y_2 \end{smallmatrix}\right], \left[\begin{smallmatrix} z_1 \\ z_2 \end{smallmatrix}\right]\right\rangle = \langle \mathbf{u}, \mathbf{w}\rangle + \alpha\langle \mathbf{v}, \mathbf{w}\rangle,\end{aligned}$$

e, portanto, a função é linear na primeira entrada. Vamos verificar ainda a simetria

$$\langle \mathbf{u}, \mathbf{v}\rangle = 2x_1y_1 - x_1y_2 - x_2y_1 + x_2y_2 = 2y_1x_1 - y_1x_2 - y_2x_1 + y_2x_2 = \langle \mathbf{v}, \mathbf{u}\rangle.$$

E, por fim,

$$\langle \mathbf{u}, \mathbf{u}\rangle = 2x_1^2 - x_1x_2 - x_2x_1 + x_2^2 = x_1^2 + x_1^2 - x_1x_2 - x_2x_1 + x_2^2 = x_1^2 + (x_1 - x_2)^2 \geq 0.$$

A expressão $x_1^2 + (x_1 - x_2)^2 = 0$ se, e somente se, $x_1 = x_2 = 0$. Portanto, essa função satisfaz a todas as propriedades requeridas para que seja um produto interno.

9.3 Normas

Lembramos que, no plano, para calcular o comprimento de um vetor $\mathbf{u} = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]$, basta calcular $\sqrt{x_1^2 + x_2^2}$. Isso se dá pela aplicação do teorema de Pitágoras. Para certificar-se de que entendeu isso, veja a figura 9.3. Falar no comprimento (ou norma) de um vetor é equivalente a calcular a distância da origem até as coordenadas que definem o vetor.

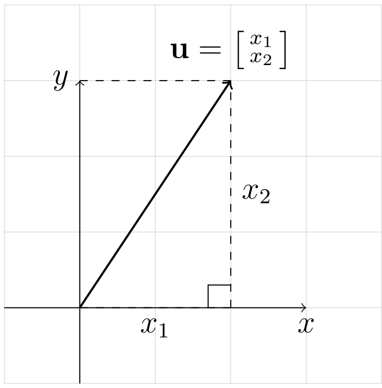


Figura 9.1: Aplicação do Teorema de Pitágoras

Observe que, se considerarmos o produto interno no plano, como o sendo o produto escalar, podemos expressar:

$$\sqrt{x_1^2 + x_2^2} = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}.$$

É possível fazer algo similar se estivermos no $\mathbb{R}^3$ com o produto interno sendo o produto escalar (repita o raciocínio acima).

Podemos generalizar o conceito de comprimento de vetor sempre que tivermos um produto interno definido, uma vez que, pela condição I_3, o produto interno de um vetor por si mesmo é sempre não negativo e, por isto, podemos definir:

Definição 9.5
Sejam $\mathbb{R}^n$ com um produto interno $\langle , \rangle$ e $\mathbf{u} \in \mathbb{R}^n$. Definimos a norma de um vetor $\mathbf{u}$ por ser

$$\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}.$$

Se $\|\mathbf{u}\| = 1$ ou, de maneira equivalente, $\langle \mathbf{u}, \mathbf{u} \rangle = 1$, dizemos que o vetor é **unitário** e que o vetor está *normalizado*. No caso em que $\mathbf{u} \neq \mathbf{0}$, com $\|\mathbf{u}\| \neq 1$, então podemos definir o **versor** desse vetor por fazer

$$\mathbf{u}' = \frac{1}{\|\mathbf{u}\|}\mathbf{u}.$$

Observe que o versor $\mathbf{u}'$ tem norma igual a 1 e a mesma direção que o vetor $\mathbf{u}$. Esse processo também é conhecido por normalização do vetor $\mathbf{u}$.

Observação 9.6
Esta observação nos será muito útil no futuro. Sejam $\mathbf{u}$ e $\mathbf{v}$ dois vetores, então:

$$\begin{aligned}\|\mathbf{u}+\mathbf{v}\|^2 &= \langle \mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{u}\rangle + \langle \mathbf{u}, \mathbf{v}\rangle + \langle \mathbf{v}, \mathbf{u}\rangle + \langle \mathbf{v}, \mathbf{v}\rangle \\ &= \|\mathbf{u}\|^2 + 2\langle \mathbf{u}, \mathbf{v}\rangle + \|\mathbf{v}\|^2 .\end{aligned}$$

Exemplo 9.7
Considere $\mathbf{u} = \left[\begin{smallmatrix} 1 \\ 2 \\ -3 \end{smallmatrix}\right] \in \mathbb{R}^3$ e o produto interno, o produto escalar. Logo a norma é

$$\|\mathbf{u}\| = \sqrt{1 + 2^2 + (-3)^2} = \sqrt{14}.$$

Observação 9.8
Se $\mathbb{R}^n$ está munido de um produto interno, então a norma satisfará à seguinte propriedade: para quaisquer $\mathbf{u}, \mathbf{v} \in V$, temos

$$\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\| \text{ (Desigualdade Triangular).}$$

Essa propriedade nos diz que o comprimento de um dos lados de um triângulo é sempre menor ou igual à soma dos comprimentos dos outros dois lados. Veja figura 9.2.

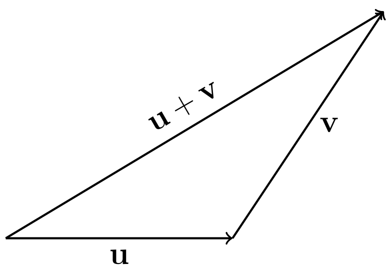


Figura 9.2: Desigualdade Triangular

Definição 9.9
Seja $\mathbb{R}^n$ com um produto interno $\langle\, , \,\rangle$. A distância entre $\mathbf{u}$ e $\mathbf{v}$ é definida por

$$\operatorname{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| .$$

Exemplo 9.10
Considerando $\mathbb{R}^2$ e o produto interno é o usual, calcule a distância entre $\mathbf{u} = \left[\begin{smallmatrix} 1 \\ -2 \end{smallmatrix}\right]$ e $\mathbf{v} = \left[\begin{smallmatrix} 3 \\ 1 \end{smallmatrix}\right]$.

$$\operatorname{dist}(\mathbf{u}, \mathbf{v}) = \left\| \left[\begin{smallmatrix} 1 \\ -2 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 3 \\ 1 \end{smallmatrix}\right] \right\| = \sqrt{\left\langle \left[\begin{smallmatrix} -2 \\ -3 \end{smallmatrix}\right], \left[\begin{smallmatrix} -2 \\ -3 \end{smallmatrix}\right] \right\rangle} = \sqrt{13} \text{ unidades.}$$

9.4 Ortogonalidade

Vamos justificar a ideia de $\mathbf{u}$ ser ortogonal ou perpendicular a $\mathbf{v}$.

Seja novamente $\mathbb{R}^2$ com o produto interno o produto escalar. Considere dois vetores não nulos e não múltiplos um do outro, $\mathbf{u} = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]$ e $\mathbf{v} = \left[\begin{smallmatrix} y_1 \\ y_2 \end{smallmatrix}\right]$. Então, podemos considerar os vetores $\mathbf{u}, \mathbf{v}, \mathbf{u} - \mathbf{v}$, são os lados de um triângulo (veja figura 9.3), e relacionar o comprimento de seus lados, usando a lei dos cossenos, se θ é o ângulo formado entre os vetores $\mathbf{u}$ e $\mathbf{v}$ então

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\,\|\mathbf{u}\|\,\|\mathbf{v}\| \cos\theta.$$

Isso é equivalente a dizer que

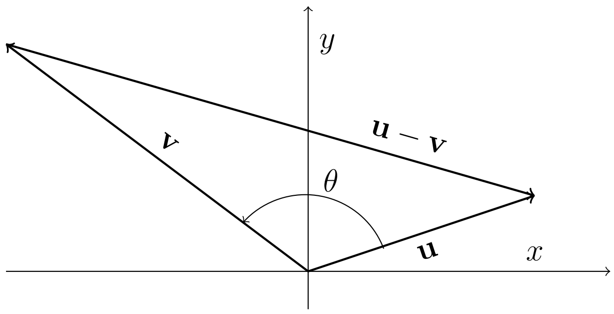


Figura 9.3: Lei dos Cossenos

$$\begin{aligned}
\|\mathbf{u}\|\,\|\mathbf{v}\| \cos\theta &= \frac{1}{2}\left[\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - \|\mathbf{u} - \mathbf{v}\|^2\right] \\
&= \frac{1}{2}\left[x_1^2 + x_2^2 + y_1^2 + y_2^2 - (x_1 - y_1)^2 - (x_2 - y_2)^2\right] \\
&= x_1 y_1 + x_2 y_2 \\
&= \langle \mathbf{u}, \mathbf{v} \rangle,
\end{aligned}$$

e temos a seguinte relação:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \|\mathbf{u}\|\,\|\mathbf{v}\| \cos\theta.$$

Logo, se $\theta = \pi/2$, isto é, $\mathbf{u}$ é ortogonal a $\mathbf{v}$, temos, neste caso, $\langle \mathbf{u}, \mathbf{v} \rangle = 0$. Reciprocamente, se $\mathbf{u}$ e $\mathbf{v}$ são dois vetores não nulos e $\langle \mathbf{u}, \mathbf{v} \rangle = 0$, então o ângulo entre eles é de $\cos\theta = 0$, portanto, $\theta = \pi/2$.

Portanto, definimos para qualquer produto interno.

Definição 9.11
Considerando o $\mathbb{R}^n$ com um produto interno, dizemos que dois vetores $\mathbf{u}$ e $\mathbf{v}$ em $\mathbb{R}^n$ são **ortogonais** (ou perpendiculares) se, e somente se, $\langle \mathbf{u}, \mathbf{v} \rangle = 0$. Neste caso, denotamos por $\mathbf{u} \perp \mathbf{v}$.

As seguintes propriedades seguem imediatamente a definição de ortogonalidade.

1) Qualquer que seja o vetor $\mathbf{u} \in \mathbb{R}^n$, temos $\mathbf{u} \perp \mathbf{0}$.

2) $\mathbf{u} \perp \mathbf{w}$ se, e somente se, $\mathbf{w} \perp \mathbf{u}$.

3) Se $\mathbf{u} \perp \mathbf{w}$, para todo $\mathbf{w} \in V$, então $\mathbf{u} = \mathbf{0}$.

4) Se $\mathbf{u} \perp \mathbf{w}$ e $\mathbf{v} \perp \mathbf{w}$, então $(\mathbf{u} + \lambda \mathbf{v}) \perp \mathbf{w}$.

Dizemos que $X = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ é um **conjunto ortogonal**, se $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0$, para todo $i \neq j$. E, se além disso, $\langle \mathbf{u}_i, \mathbf{u}_i \rangle = 1$, para todo $i = 1, 2, \ldots, n$, dizemos que X é **ortonormal**.

Teorema 9.12
Considere o $\mathbb{R}^n$ munido de produto interno $\langle\, , \rangle$. Se X é um conjunto ortogonal formado por vetores não nulos, então X é LI.

Demonstração: Veja o exercício R9.1. □

Exemplo 9.13
Considere $\mathbb{R}^3$ e o produto interno $\langle\, , \rangle$ o produto escalar do $\mathbb{R}^3$ e $X = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix} \right\}$. Vamos verificar que esse conjunto é ortogonal. Para isso, considere

$$\left\langle \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \right\rangle = -1 + 1 = 0, \quad \left\langle \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix} \right\rangle = -1 - 1 + 2 = 0$$

e

$\left\langle \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix} \right\rangle = 1 - 1 = 0.$ Portanto, esse conjunto é ortogonal.

Observação 9.14
Teorema de Pitágoras: Sejam $\mathbb{R}^n$ com um produto interno e $\mathbf{u}$ e $\mathbf{v}$ dois vetores. Então, sabemos que

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + 2\langle \mathbf{u}, \mathbf{v} \rangle + \|\mathbf{v}\|^2.$$

Se $\mathbf{u} \perp \mathbf{v}$, temos que $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ e daí

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2,$$

que é o Teorema de Pitágoras para o caso vetorial.

9.4.1 Complemento Ortogonal

Seja F um subconjunto do $\mathbb{R}^n$, o qual está munido de um produto interno $\langle\, , \rangle$. O complemento ortogonal de F, denotado por $F^\perp$, consiste nos vetores de $\mathbb{R}^n$ que são ortogonais a cada vetor $\mathbf{v} \in F$, ou seja,

$$F^\perp = \{\mathbf{u} \in \mathbb{R}^n : \langle \mathbf{u}, \mathbf{v} \rangle = 0, \text{ para todo } \mathbf{v} \in F\}.$$

No caso em que $F = \{\mathbf{v}\}$, temos que

$$v^\perp = \{\mathbf{u} \in \mathbb{R}^n : \langle \mathbf{u}, \mathbf{v} \rangle = 0\}$$

É fácil perceber que nesta situação $v^\perp$ é um subespaço vetorial.

Exemplo 9.15
Considere $F = \left\{ \mathbf{u} = \left[\begin{smallmatrix} 1 \\ 2 \\ -1 \end{smallmatrix}\right], \mathbf{v} = \left[\begin{smallmatrix} -2 \\ 0 \\ 1 \end{smallmatrix}\right] \right\}$ um subconjunto de $\mathbb{R}^3$. Encontre $F^\perp$.

Queremos determinar $\mathbf{w} = \left[\begin{smallmatrix} x \\ y \\ z \end{smallmatrix}\right]$, tais que $\langle \mathbf{w}, \mathbf{u} \rangle = 0$ e, também, $\langle \mathbf{w}, \mathbf{v} \rangle = 0$. Fazendo as contas, obtemos

$$\begin{cases} x + 2y - z = 0 \\ \quad -2x + z = 0 \end{cases} \quad \Leftrightarrow \quad \begin{cases} y = 3/2\, x \\ z = 2x. \end{cases}$$

Portanto, $F^\perp = \left\{ \left[\begin{smallmatrix} x \\ 3/2\, x \\ 2x \end{smallmatrix}\right] : x \in \mathbb{R} \right\} = \operatorname{Span}\left\{ \left[\begin{smallmatrix} 1 \\ 3/2 \\ 2 \end{smallmatrix}\right] \right\}$.

Teorema 9.16
Seja F um subconjunto do $\mathbb{R}^n$, o qual está munido de um produto interno, então $F^\perp$ é um subespaço vetorial de $\mathbb{R}^n$.

Demonstração: Veja o exercício R9.4. $\square$

9.5 Projeções Ortogonais

O objetivo é discutir como definir operadores lineares que projetam ortogonalmente um vetor sobre um subespaço vetorial W do $\mathbb{R}^n$.

Inicialmente, considere o vetor unitário $\mathbf{v}$ e outro vetor $\mathbf{u}$ qualquer do $\mathbb{R}^n$. Chamamos o vetor $\langle \mathbf{u}, \mathbf{v} \rangle \mathbf{v}$ de projeção ortogonal de $\mathbf{u}$ sobre $\mathbf{v}$. Veja a figura 9.4.

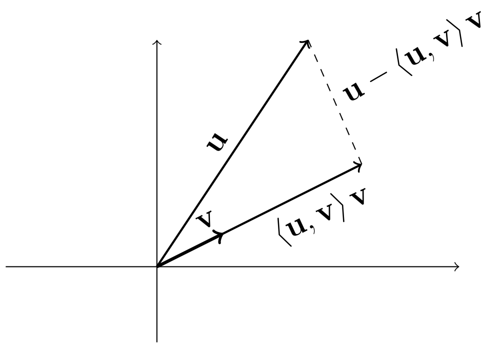


Figura 9.4: Projeção ortogonal sobre o $\mathbf{v}$

Vamos justificar esse nome, verificando que $\mathbf{w} = \mathbf{u} - \langle \mathbf{u}, \mathbf{v} \rangle \mathbf{v}$ é perpendicular ao vetor $\mathbf{v}$.

$$\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{u} - \langle \mathbf{u}, \mathbf{v} \rangle \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle - \langle \mathbf{u}, \mathbf{v} \rangle \langle \mathbf{v}, \mathbf{v} \rangle = 0,$$

pois $\langle \mathbf{v}, \mathbf{v} \rangle = 1$. E, no caso geral, se $\mathbf{v}$ é um vetor não nulo qualquer, consideremos o seu versor, $\mathbf{v}' = \frac{1}{\|\mathbf{v}\|}\mathbf{v}$. Então, novamente $\langle \mathbf{u}, \mathbf{v}' \rangle \mathbf{v}'$ é a projeção ortogonal de $\mathbf{u}$ sobre $\mathbf{v}'$ e, substituindo $\mathbf{v}'$ por $\mathbf{v}$, temos:

$$\langle \mathbf{u}, \mathbf{v}' \rangle \mathbf{v}' = \left\langle \mathbf{u}, \frac{1}{\|\mathbf{v}\|}\mathbf{v} \right\rangle \frac{1}{\|\mathbf{v}\|}\mathbf{v} = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{v}\|^2}\mathbf{v} = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}\mathbf{v}.$$

Denotamos a projeção ortogonal de $\mathbf{u}$ sobre $\mathbf{v}$ por $\operatorname{proj}_{\mathbf{v}}(\mathbf{u}) = \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}\mathbf{v}$.

Observe ainda que a norma do vetor $\mathbf{v}$ é indiferente para o resultado. Por isso, é mais interessante considerarmos o subespaço vetorial gerado por $\mathbf{v}$, isto é, $W = \operatorname{Span}\{\mathbf{v}\}$, e definirmos

$$\operatorname{proj}_W(\mathbf{u}) = \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}\mathbf{v}, \text{ onde } \mathbf{0} \neq \mathbf{v} \in W.$$

Antes de continuarmos, vamos dar um exemplo da forma que esse operador assume no plano.

Exemplo 9.17

Estamos interessados em determinar as fórmulas de uma transformação linear que projeta um dado vetor $\mathbf{u} \in \mathbb{R}^2$ sobre uma reta, passando pela origem, e que, portanto, tem equação $y = ax$ para algum $a \in R$ fixo.

Logo, a reta é gerada pelo vetor $\mathbf{v} = \left[\begin{smallmatrix}1\\a\end{smallmatrix}\right]$, isto é, qualquer ponto da reta é um múltiplo deste vetor. Considere $\mathbf{u} = \left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]$. Queremos encontrar $\lambda \in \mathbb{R}$, tal que $\lambda\mathbf{v}$ seja a projeção ortogonal de $\mathbf{u}$ sobre a reta $y = ax$. Como vimos anteriormente, outra maneira de expressar essa relação, é pedir que os vetores $\mathbf{v}$ e $\mathbf{u} - \lambda\mathbf{v}$ sejam perpendiculares, isto é,

$$\langle \mathbf{v}, \mathbf{u} - \lambda\mathbf{v}\rangle = \langle \mathbf{v},\mathbf{u}\rangle - \lambda \langle \mathbf{v},\mathbf{v}\rangle = 0.$$

Isolando λ, temos que

$$\lambda = \frac{\langle \mathbf{v},\mathbf{u}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle} = \frac{\left\langle \left[\begin{smallmatrix}1\\a\end{smallmatrix}\right], \left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]\right\rangle}{\left\langle \left[\begin{smallmatrix}1\\a\end{smallmatrix}\right], \left[\begin{smallmatrix}1\\a\end{smallmatrix}\right]\right\rangle} = \frac{x + ay}{1 + a^2}.$$

Portanto, se definirmos $P : \mathbb{R}^2 \to \mathbb{R}^2$ por $\mathbf{u} \mapsto \operatorname{proj}_{\mathbf{v}}(\mathbf{u}) = \lambda\mathbf{v}$, este operador terá a seguinte expressão:

$$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} \left(\frac{1}{1+a^2}\right) x + \left(\frac{a}{1+a^2}\right) y \\ \left(\frac{a}{1+a^2}\right) x + \left(\frac{a^2}{1+a^2}\right) y \end{bmatrix}.$$

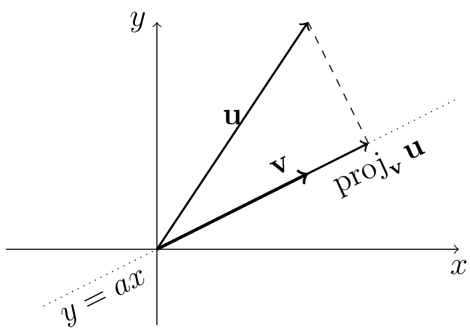


Figura 9.5: Projeção

No exemplo 9.18, veremos que a reflexão de um vetor, em torno de uma reta que passa pela origem, está intimamente ligada com a projeção do vetor sobre esta mesma reta.

Exemplo 9.18
A reflexão de um vetor $\mathbf{u}$ sobre uma reta $y = ax$ nos dá um vetor $S\mathbf{u}$ do lado oposto da reta, de tal maneira que o segmento de reta determinado por $\mathbf{u}$ e $S\mathbf{u}$ é perpendicular a reta $y = ax$ que corta este segmento no ponto médio, conforme figura 9.6.

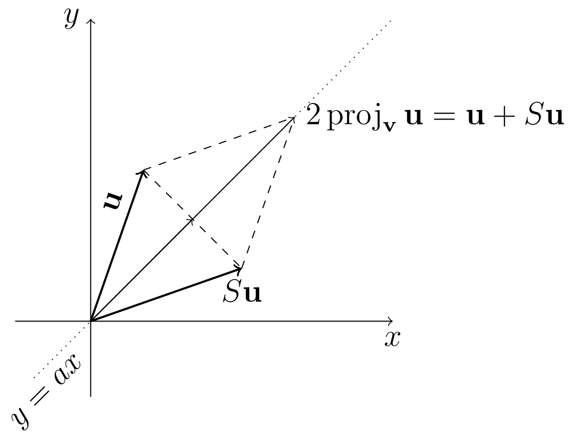


Figura 9.6: Reflexão

Observe na figura 9.6 que o vetor $\mathbf{u}+S\mathbf{u}$ está sobre a reta $y = ax$, uma vez que esta reta é exatamente a diagonal do paralelogramo determinado pelos vetores $\mathbf{u}$ e $S\mathbf{u}$. Além disso, o vetor $\mathbf{u} + S\mathbf{u}$ é igual a duas vezes a projeção ortogonal de $\mathbf{u}$ sobre o vetor $\mathbf{v} = \left[\begin{smallmatrix} 1 \\ a \end{smallmatrix}\right]$ que determina a reta. E, portanto,

$$2\,\mathrm{proj}_{\mathbf{v}}\,\mathbf{u} = \mathbf{u} + S\mathbf{u} \Leftrightarrow S\mathbf{u} = 2\,\mathrm{proj}_{\mathbf{v}}\,\mathbf{u} - \mathbf{u}.$$

Em termos de fórmulas, obtemos:

$$S\begin{bmatrix} x \\ y \end{bmatrix} = 2\begin{bmatrix} \left(\frac{1}{1+a^2}\right)x + \left(\frac{a}{1+a^2}\right)y \\ \left(\frac{a}{1+a^2}\right)x + \left(\frac{a^2}{1+a^2}\right)y \end{bmatrix} - \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \left(\frac{1-a^2}{1+a^2}\right)x + \left(\frac{2a}{1+a^2}\right)y \\ \left(\frac{2a}{1+a^2}\right)x + \left(\frac{a^2-1}{1+a^2}\right)y \end{bmatrix}.$$

9.5.1 Desigualdade de Cauchy-Schwarz

Vamos retornar à situação $\mathbf{z} = \mathrm{proj}_{\mathbf{v}}(\mathbf{u})$ e definir $\mathbf{w} = \mathbf{u} - \mathbf{z}$, isso implica que $\mathbf{u} = \mathbf{w} + \mathbf{z}$, com $\mathbf{w} \perp \mathbf{z}$. Pelo teorema de Pitágoras, temos que: $\|\mathbf{u}\|^2 = \|\mathbf{w}\|^2 + \|\mathbf{z}\|^2$. Em particular, $\|\mathbf{z}\| \leq \|\mathbf{u}\|$, isto é, o comprimento da projeção $\mathrm{proj}_{\mathbf{v}}(\mathbf{u})$ é sempre menor ou igual ao comprimento de $\mathbf{u}$.

Mas a norma de $\mathrm{proj}_{\mathbf{v}}(\mathbf{u})$ é $\|\langle \mathbf{u}, \mathbf{v}\rangle\| / \|\mathbf{v}\|$. Segue que $\|\langle \mathbf{u}, \mathbf{v}\rangle\| / \|\mathbf{v}\| \leq \|\mathbf{u}\|$, ou seja,

$$|\langle \mathbf{u}, \mathbf{v}\rangle| \leq \|\mathbf{u}\|\,\|\mathbf{v}\| \ \text{(Desigualdade de Cauchy-Schwarz)}.$$

9.5.2 Processo de Ortonormalização de Gram-Schmidt

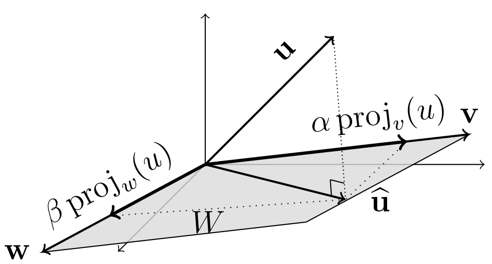
O processo de Gram-Schmidt é um procedimento que inicia com um conjunto LI $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s\}$ de vetores de $\mathbb{R}^n$ e retorna um conjunto ortonormal $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_s\}$, com a seguinte propriedade: para cada k, com $1 \leq k \leq s$, os vetores $\mathbf{u}_1$,$\mathbf{u}_2$,$\ldots$, $\mathbf{u}_k$ pertencem ao subespaço vetorial $\operatorname{Span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$.

Para entender o processo de ortogonalização de Gram-Schmidt, vamos voltar a nossa discussão de como obter a projeção ortogonal de um vetor $\mathbf{u}$ sobre um subespaço vetorial W, de dimensão dois, do $\mathbb{R}^n$. Como estamos interessados em encontrar uma fórmula para a projeção ortogonal de um vetor $\mathbf{u}$ sobre W, considere $\mathbf{v}, \mathbf{w} \in W$, tal que $W = \operatorname{Span}\{\mathbf{v}, \mathbf{w}\}$. Lembrando que já sabemos calcular $\operatorname{proj}_{\mathbf{v}} \mathbf{u}$ e $\operatorname{proj}_{\mathbf{w}} \mathbf{u}$. Então, certamente existem α e $\beta \in \mathbb{R}$, tais que o vetor $\widehat{\mathbf{u}} = \alpha \operatorname{proj}_{\mathbf{v}} \mathbf{u} + \beta \operatorname{proj}_{\mathbf{w}} \mathbf{u}$ é a projeção ortogonal de $\mathbf{u}$ sobre W. Precisamos determinar α e β. Como sabemos que $(\mathbf{u} - \widehat{\mathbf{u}}) \perp \mathbf{v}$ deve acontecer, temos:

$$\begin{aligned}
\langle (\mathbf{u} - \widehat{\mathbf{u}}), \mathbf{v} \rangle &= \langle \mathbf{u} - (\alpha \operatorname{proj}_{\mathbf{v}} \mathbf{u} + \beta \operatorname{proj}_{\mathbf{w}} \mathbf{u}), \mathbf{v} \rangle \\
&= \langle \mathbf{u}, \mathbf{v} \rangle - \alpha \langle \operatorname{proj}_{\mathbf{v}} \mathbf{u}, \mathbf{v} \rangle - \beta \langle \operatorname{proj}_{\mathbf{w}} \mathbf{u}, \mathbf{v} \rangle \\
&= \langle \mathbf{u}, \mathbf{v} \rangle - \alpha \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle} \langle \mathbf{v}, \mathbf{v} \rangle - \beta \frac{\langle \mathbf{u}, \mathbf{w} \rangle}{\langle \mathbf{w}, \mathbf{w} \rangle} \langle \mathbf{w}, \mathbf{v} \rangle \\
&= \langle \mathbf{u}, \mathbf{v} \rangle - \alpha \langle \mathbf{u}, \mathbf{v} \rangle - \beta \frac{\langle \mathbf{u}, \mathbf{w} \rangle \langle \mathbf{w}, \mathbf{v} \rangle}{\langle \mathbf{w}, \mathbf{w} \rangle} = 0.
\end{aligned}$$

E, como também $(\mathbf{u} - \widehat{\mathbf{u}}) \perp \mathbf{w}$ deve acontecer, obtemos que:

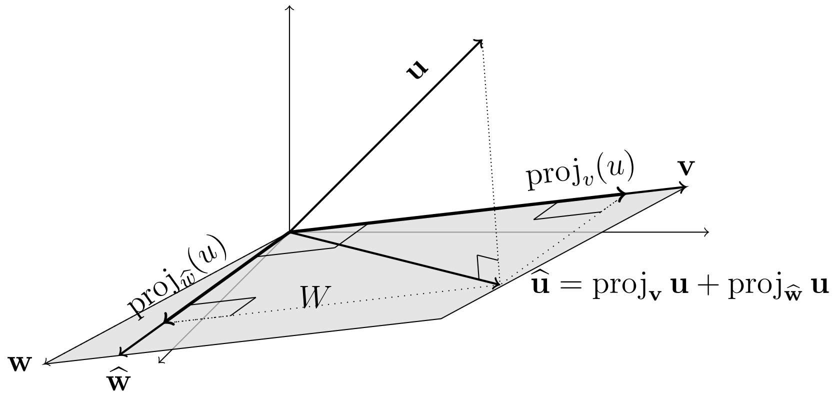
$$\begin{aligned}
\langle (\mathbf{u} - \widehat{\mathbf{u}}), \mathbf{w} \rangle &= \langle \mathbf{u} - (\alpha \operatorname{proj}_{\mathbf{v}} \mathbf{u} + \beta \operatorname{proj}_{\mathbf{w}} \mathbf{u}), \mathbf{w} \rangle \\
&= \langle \mathbf{u}, \mathbf{w} \rangle - \alpha \frac{\langle \mathbf{u}, \mathbf{v} \rangle \langle \mathbf{v}, \mathbf{w} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle} - \beta \langle \mathbf{u}, \mathbf{w} \rangle = 0.
\end{aligned}$$



Encontrar α e β é equivalente a encontrar a solução de um sistema linear com duas equações nas variáveis α e β. Mas este sistema possui uma solução muito simples se $\mathbf{v} \perp \mathbf{w}$: nesta circunstância a solução é $\alpha = 1 = \beta$. Além disso, se $\mathbf{v}$ e $\mathbf{w}$ não são perpendiculares, sem nenhuma perda, podemos escolher os vetores $\mathbf{v}, \widehat{\mathbf{w}} = \mathbf{w} - \operatorname{proj}_{\mathbf{v}} \mathbf{w}$, e teremos que $\mathbf{v} \perp \widehat{\mathbf{w}}$, uma vez que ainda $W = \operatorname{Span}\{\mathbf{v}, \widehat{\mathbf{w}}\}$.

Juntando tudo isso, podemos definir o operador projeção ortogonal de $\mathbf{u}$ sobre W por ser

$$\operatorname{proj}_W \mathbf{u} = \operatorname{proj}_{\mathbf{v}} \mathbf{u} + \operatorname{proj}_{\widehat{\mathbf{w}}} \mathbf{u}.$$



Essa situação se generaliza para espaços em qualquer dimensão, e é a base do processo de ortonormalização de Gram-Schmidt que descrevemos abaixo. O processo inicia com um conjunto LI $X = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s\}$ de vetores de $\mathbb{R}^n$ e retorna um conjunto ortonormal $Y = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_s\}$, com a seguinte propriedade: para cada k, com $1 \leq k \leq s$, os vetores $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ pertencem ao subespaço vetorial $\operatorname{Span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$.

Vamos descrever o procedimento, fazendo

$$
\begin{aligned}
&\mathbf{v}_1' = \mathbf{v}_1 \text{ e } \{\mathbf{u}_1'\} \subset \operatorname{Span}\{\mathbf{v}_1\}\\
&\mathbf{v}_2' = \mathbf{v}_2 - \frac{\langle \mathbf{v}_2, \mathbf{v}_1' \rangle}{\langle \mathbf{v}_1', \mathbf{v}_1' \rangle}\mathbf{v}_1' \text{ e } \{\mathbf{v}_1', \mathbf{v}_2'\} \subset \operatorname{Span}\{\mathbf{v}_1, \mathbf{v}_2\}\\
&\mathbf{v}_3' = \mathbf{v}_3 - \left(\frac{\langle \mathbf{v}_3, \mathbf{v}_1' \rangle}{\langle \mathbf{v}_1', \mathbf{v}_1' \rangle}\mathbf{v}_1' + \frac{\langle \mathbf{v}_3, \mathbf{v}_2' \rangle}{\langle \mathbf{v}_2', \mathbf{v}_2' \rangle}\mathbf{v}_2'\right) \text{ e } \{\mathbf{v}_1', \mathbf{v}_2', \mathbf{v}_3'\} \subset \operatorname{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}\\
&\cdots\cdots\cdots\\
&\mathbf{v}_k' = \mathbf{v}_k - \left(\sum_{j=1}^{k} \frac{\langle \mathbf{v}_k, \mathbf{v}_j' \rangle}{\langle \mathbf{v}_j', \mathbf{v}_j' \rangle}\mathbf{v}_j'\right) \text{ e } \{\mathbf{v}_1', \mathbf{v}_2', \ldots, \mathbf{v}_k'\} \subset \operatorname{Span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}\\
&\cdots\cdots\cdots\\
&\mathbf{v}_s' = \mathbf{v}_s - \left(\sum_{j=1}^{s} \frac{\langle \mathbf{v}_s, \mathbf{v}_j' \rangle}{\langle \mathbf{v}_j', \mathbf{v}_j' \rangle}\mathbf{v}_j'\right) \text{ e } \{\mathbf{v}_1', \mathbf{v}_2', \ldots, \mathbf{v}_s'\} \subset \operatorname{Span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s\}.
\end{aligned}
$$

E, finalmente, vamos normalizar os vetores, isto é,

$$\mathbf{u}_1 = \frac{1}{\|\mathbf{v}_1'\|}\mathbf{v}_1', \ldots, \mathbf{u}_s = \frac{1}{\|\mathbf{v}_s'\|}\mathbf{v}_s'.$$

Exemplo 9.19

Seja o $\mathbb{R}^3$ com o produto escalar, considere a seguinte base $\left\{\mathbf{v}_1 = \begin{bmatrix}1\\1\\1\end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix}0\\2\\1\end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix}0\\0\\3\end{bmatrix}\right\}$ do $\mathbb{R}^3$ e aplique o processo de ortonormalização de Gram-Schmidt, para obtermos uma base ortogonal:

$$\mathbf{v}_1' = \mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\mathbf{v}_2' = \mathbf{v}_2 - \frac{\langle \mathbf{v}_2, \mathbf{v}_1' \rangle}{\langle \mathbf{v}_1', \mathbf{v}_1' \rangle}\mathbf{v}_1' = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} - \frac{\left\langle \left[\begin{smallmatrix} 0 \\ 2 \\ 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right] \right\rangle}{\left\langle \left[\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right] \right\rangle} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} - \frac{3}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

$$\mathbf{v}_3' = \mathbf{v}_3 - \left(\frac{\langle \mathbf{v}_3, \mathbf{v}_1' \rangle}{\langle \mathbf{v}_1', \mathbf{v}_1' \rangle}\mathbf{v}_1' + \frac{\langle \mathbf{v}_3, \mathbf{v}_2' \rangle}{\langle \mathbf{v}_2', \mathbf{v}_2' \rangle}\mathbf{v}_2' \right)$$

$$= \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix} - \left(\frac{\left\langle \left[\begin{smallmatrix} 0 \\ 0 \\ 3 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right] \right\rangle}{\left\langle \left[\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right] \right\rangle} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \frac{\left\langle \left[\begin{smallmatrix} 0 \\ 0 \\ 3 \end{smallmatrix}\right], \left[\begin{smallmatrix} -1 \\ 1 \\ 0 \end{smallmatrix}\right] \right\rangle}{\left\langle \left[\begin{smallmatrix} -1 \\ 1 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} -1 \\ 1 \\ 0 \end{smallmatrix}\right] \right\rangle} \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \right)$$

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

Normalizando, temos:

$$\mathbf{u}_1 = \frac{1}{\|\mathbf{v}_1'\|}\mathbf{v}_1' = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}; \ \mathbf{u}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \ \text{e} \ \mathbf{u}_3 = \frac{1}{\sqrt{6}}\begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix},$$

a qual é uma base ortonormal do $\mathbb{R}^3$.

Observação 9.20
Como consequência do processo de ortonormalização de Gram-Schmidt, temos que todo espaço vetorial de dimensão finita, munido de um produto interno, admite uma base ortonormal. De fato, seja $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ uma base qualquer de $\mathbb{R}^n$, aplicando o processo de Gram-Schmidt, obtemos $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ ortonormal, que gera $\mathbb{R}^n$. Pelo teorema 9.12, sabemos que $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ é LI e, portanto, ele é uma base de $\mathbb{R}^n$.

Observação 9.21
Uma outra consequência do conceito de projeção ortogonal é que o mesmo minimiza a distância de um vetor $\mathbf{u} \in \mathbb{R}^n$ a um subespaço vetorial $W \subset \mathbb{R}^n$. Para compreender bem esta propriedade, suponha que $\widehat{\mathbf{u}} = \operatorname{proj}_W(\mathbf{u})$, isto é, $\widehat{\mathbf{u}}$ é a projeção ortogonal de $\mathbf{u}$ sobre o subespaço vetorial W. Vamos mostrar que

$$\|\mathbf{u} - \widehat{\mathbf{u}}\| \leq \|\mathbf{u} - \mathbf{v}\|, \text{ para todo } \mathbf{v} \in W.$$

Seja $\mathbf{v}$ um vetor qualquer de W, então $\widehat{\mathbf{u}} - \mathbf{v} \in W$ e $(\widehat{\mathbf{u}} - \mathbf{v}) \perp (\mathbf{u} - \widehat{\mathbf{u}})$, mas pelo teorema de Pitágoras, para espaços vetoriais,

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u} - \widehat{\mathbf{u}} + \widehat{\mathbf{u}} - \mathbf{v}\|^2 = \|\mathbf{u} - \widehat{\mathbf{u}}\|^2 + \|\widehat{\mathbf{u}} - \mathbf{v}\|^2.$$

Portanto, $\|\mathbf{u} - \mathbf{v}\| \geq \|\mathbf{u} - \widehat{\mathbf{u}}\|$, para todo $\mathbf{v} \in W$.

Essa propriedade possui numerosas aplicações, tais como o **Método dos Mínimos Quadrados** ou **Regressão linear, quadrática ou Regressão exponencial**, como é conhecido na Estatística.

9.6 Matrizes e Operadores Ortogonais

As matrizes ortogonais são interessantes, pois com elas podemos executar movimentos sem deformação no $\mathbb{R}^n$.

Vamos fazer uma observação que serve como motivação para a definição da matriz ortogonal. Seja $V = \mathbb{R}^3$ e o produto interno o produto escalar do $\mathbb{R}^3$. Se $\left\{\mathbf{u}_1 = \left[\begin{smallmatrix} x_1 \\ x_2 \\ x_3 \end{smallmatrix}\right], \mathbf{u}_2 = \left[\begin{smallmatrix} y_1 \\ y_2 \\ y_3 \end{smallmatrix}\right], \mathbf{u}_3 = \left[\begin{smallmatrix} z_1 \\ z_2 \\ z_3 \end{smallmatrix}\right]\right\}$ é uma base ortonormal, isto é, $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \mathbf{u}_i^t \mathbf{u}_j = 0$, se $i \neq j$, se $i = j$, então $\langle \mathbf{u}_i, \mathbf{u}_i \rangle = \mathbf{u}_i^t \mathbf{u}_i = \|\mathbf{u}_i\|^2 = 1$.

Logo, se montarmos a matriz A, colocando os vetores $\mathbf{u}_i$ nas colunas, acharemos $A = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$, e ao calcularmos

$$A^t A = \begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{bmatrix} \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{bmatrix} = \begin{bmatrix} \langle \mathbf{u}_1, \mathbf{u}_1 \rangle & \langle \mathbf{u}_1, \mathbf{u}_2 \rangle & \langle \mathbf{u}_1, \mathbf{u}_3 \rangle \\ \langle \mathbf{u}_2, \mathbf{u}_1 \rangle & \langle \mathbf{u}_2, \mathbf{u}_2 \rangle & \langle \mathbf{u}_2, \mathbf{u}_3 \rangle \\ \langle \mathbf{u}_3, \mathbf{u}_1 \rangle & \langle \mathbf{u}_3, \mathbf{u}_2 \rangle & \langle \mathbf{u}_3, \mathbf{u}_3 \rangle \end{bmatrix}.$$

E, usando que $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ é uma base ortonormal, teremos que $A^t A = I$ (veja definição 9.11).

Definição 9.22
Dizemos que uma matriz quadrada A é uma matriz ortogonal se $A^t A = I$.

Exemplo 9.23
De acordo com o exemplo 9.19, o conjunto

$$\left\{ \mathbf{u}_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{u}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{u}_3 = \frac{1}{\sqrt{6}} \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix} \right\}$$

é uma base ortonormal e, portanto, a matriz

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

é ortogonal.

Teorema 9.24
Seja $\langle \, , \, \rangle$ um produto interno em $\mathbb{R}^n$, considere α e β, duas bases ortonormais em $\mathbb{R}^n$. Então a matriz de mudança de coordenadas $[I]^\alpha_\beta$ é uma matriz ortogonal.

Demonstração: Vamos fazer a demonstração para o caso em que $n = 3$. Acreditamos que você será capaz de dar uma prova para o caso geral. Sejam $\alpha = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ e $\beta = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ duas bases ortonormais de $\mathbb{R}^3$. Para achar a matriz $[I]^\alpha_\beta$, precisamos encontrar as coordenadas dos vetores da base β com respeito à base α. Digamos que

$$\begin{aligned} \mathbf{u}_1 &= a_{11}\mathbf{v}_1 + a_{21}\mathbf{v}_2 + a_{31}\mathbf{v}_3 \\ \mathbf{u}_2 &= a_{12}\mathbf{v}_1 + a_{22}\mathbf{v}_2 + a_{32}\mathbf{v}_3 \\ \mathbf{u}_3 &= a_{13}\mathbf{v}_1 + a_{23}\mathbf{v}_2 + a_{33}\mathbf{v}_3. \end{aligned}$$

A matriz fica:

$$[I]^{\alpha}_{\beta} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

Queremos ver que essa matriz é ortogonal. Para isso, observe que, como α e β são ortonormais, obtemos:

$$\begin{aligned} 1 = \langle u_i, u_i \rangle &= \langle a_{1i}\mathbf{v}_1 + a_{2i}\mathbf{v}_2 + a_{3i}\mathbf{v}_3, a_{1i}\mathbf{v}_1 + a_{2i}\mathbf{v}_2 + a_{3i}\mathbf{v}_3 \rangle \\ &= a_{1i}^2 + a_{2i}^2 + a_{3i}^2, \end{aligned}$$

e, também, se $i \neq j$, temos:

$$\begin{aligned} 0 = \langle u_i, u_j \rangle &= \langle a_{1i}\mathbf{v}_1 + a_{2i}\mathbf{v}_2 + a_{3i}\mathbf{v}_3, a_{1j}\mathbf{v}_1 + a_{2j}\mathbf{v}_2 + a_{3j}\mathbf{v}_3 \rangle \\ &= a_{1i}a_{1j} + a_{2i}a_{2j} + a_{3i}a_{3j}. \end{aligned}$$

Isso mostra que $[I]^{\alpha}_{\beta}$ é ortogonal. □

Vamos definir agora os operadores ortogonais que estão conectados com as matrizes ortogonais

Definição 9.25
Sejam $\langle\ ,\ \rangle$ um produto interno de $\mathbb{R}^n$ e $T : \mathbb{R}^n \to \mathbb{R}^n$ um operador linear, se a matriz $[T]^{\alpha}_{\alpha}$, com respeito a alguma base ortonormal α, for ortogonal, dizemos que T é um operador ortogonal.

Exemplo 9.26
Quando tratamos de rotações de um ângulo θ em torno da origem, ao calcular a matriz deste operador linear em termos da base canônica, obtivemos, no exemplo 5.8, a matriz

$$A = \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix}.$$

Agora, se temos os vetores $\mathbf{u} = \begin{bmatrix} \cos\theta \\ \operatorname{sen}\theta \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} -\operatorname{sen}\theta \\ \cos\theta \end{bmatrix}$, obtidos por tomar as colunas da matriz A, veremos que $\langle \mathbf{u}, \mathbf{u} \rangle = \cos^2\theta + \operatorname{sen}^2\theta = 1 = \langle \mathbf{v}, \mathbf{v} \rangle$. Além disso, $\langle \mathbf{u}, \mathbf{v} \rangle = 0$, e $\{\mathbf{u}, \mathbf{v}\}$ é uma base ortonormal. Portanto, A é uma matriz ortogonal.

Teorema 9.27
Sejam $\langle\ ,\ \rangle$ um produto interno em $\mathbb{R}^n$ e $T : \mathbb{R}^n \to \mathbb{R}^n$ um operador linear, então as seguintes condições são equivalentes:

(1) *T é um operador ortogonal;*

(2) *A matriz de T com respeito a qualquer base ortonormal é ortogonal;*

(3) $\langle T(\mathbf{u}), T(\mathbf{v}) \rangle = \langle \mathbf{u}, \mathbf{v} \rangle, \quad \forall\, \mathbf{u}$ *e* $\mathbf{v} \in V$.

Exercícios resolvidos

R9.1. Prove o teorema 9.12: Seja $\langle , \rangle$ um produto interno em $\mathbb{R}^n$. Se X é um conjunto ortogonal formado por vetores não nulos, então X é LI.

Solução: Considere $X = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ um conjunto formado por vetores não nulo e ortogonais. Queremos verificar que a única solução do sistema

$$\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \cdots + \alpha_n\mathbf{u}_n = 0$$

é a solução trivial. Se fizermos o produto interno com $\mathbf{u}_i$, obtemos

$$\langle \alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \cdots + \alpha_n\mathbf{u}_n, \mathbf{u}_i \rangle = \langle 0, \mathbf{u}_i \rangle = 0.$$

Logo, $\alpha_i \langle \mathbf{u}_i, \mathbf{u}_i \rangle = 0 \Rightarrow \alpha_i = 0$, para todo $i = 1, 2, \ldots, n$ e, portanto, a única solução é a trivial $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$. Por isso, X é LI. $\square$

R9.2. Prove o teorema 9.27: Sejam $\langle \, , \, \rangle$ um produto interno em $\mathbb{R}^n$ e $T : \mathbb{R}^n \to \mathbb{R}^n$ um operador linear, então as seguintes condições são equivalentes:

(1) T é um operador ortogonal;

(2) A matriz de T com respeito a qualquer base ortonormal é ortogonal;

(3) $\langle T(\mathbf{u}), T(\mathbf{v}) \rangle = \langle \mathbf{u}, \mathbf{v} \rangle$, $\quad \forall\, \mathbf{u}$ e $\mathbf{v} \in V$.

Solução: (1) $\Rightarrow$ (2) Esta implicação segue da seguinte propriedade: digamos que A e B sejam matrizes ortogonais, então AB também é uma matriz ortogonal. De fato, se calcularmos $(AB)^tAB = B^t(A^tA)B = B^tIB = B^tB = I$. Vamos aplicar esta propriedade para provar a implicação. Iniciemos com $\alpha \subset V$, uma base ortonormal, na qual $[T]^\alpha_\alpha$ é ortonormal, e $\beta \subset V$, uma outra base ortonormal qualquer. Podemos calcular as matrizes de mudança de coordenadas $[I]^\alpha_\beta$ e $[I]^\beta_\alpha$, que são ortogonais pelo lema 9.24. Então, como

$$[T]^\beta_\beta = [I]^\alpha_\beta [T]^\alpha_\alpha [I]^\beta_\alpha,$$

segue da observação inicial que $[T]^\beta_\beta$ também é ortogonal.

(2) $\Rightarrow$ (3) Seja $\beta = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ uma base ortonormal de V e seja

$$A = [T]^\beta_\beta = [a_{ij}].$$

Então, os escalares a_{ij} são obtidos por

$$T(\mathbf{v}_j) = \sum_{i=1}^{n} a_{ij}\mathbf{v}_i \text{ e } T(\mathbf{v}_k) = \sum_{r=1}^{n} a_{rk}\mathbf{v}_r, \text{ com } j, k = 1, \ldots, n.$$

E temos

$$\begin{aligned}
\langle T(\mathbf{v}_j), T(\mathbf{v}_k)\rangle &= \left\langle \sum_{i=1}^{n} a_{ij}\mathbf{v}_i, \sum_{r=1}^{n} a_{rk}\mathbf{v}_r \right\rangle \\
&= \sum_{i=1}^{n}\sum_{r=1}^{n} a_{ij}a_{rk}\left\langle \mathbf{v}_i, \mathbf{v}_r\right\rangle \\
&= \sum_{i=1}^{n} a_{ij}a_{ik} \text{ e como a matriz } [a_{ij}] \text{ é ortogonal} \\
&= \delta_{jk} \\
&= \langle \mathbf{v}_j, \mathbf{v}_k\rangle
\end{aligned}$$

E $\langle T(\mathbf{v}_j), T(\mathbf{v}_k)\rangle = \langle \mathbf{v}_j, \mathbf{v}_k\rangle$ e a propriedade é válida para os vetores da base inicial. Mas então, se $\mathbf{u}$ e $\mathbf{v}$ são vetores quaisquer de V, então podemos escrever $\mathbf{u} = \sum_{j=1}^{n} x_j\mathbf{v}_j$ e $\mathbf{v} = \sum_{k=1}^{n} x_k\mathbf{v}_k$, logo,

$$\begin{aligned}
\langle T(\mathbf{u}), T(\mathbf{v})\rangle &= \left\langle \sum_{j=1}^{n} x_j T(\mathbf{v}_j), \sum_{k=1}^{n} y_k T(\mathbf{v}_k) \right\rangle \\
&= \sum_{j=1}^{n}\sum_{k=1}^{n} x_j y_k \left\langle T(\mathbf{v}_j), T(\mathbf{v}_k)\right\rangle \\
&= \sum_{j=1}^{n}\sum_{k=1}^{n} x_j y_k \left\langle \mathbf{v}_j, \mathbf{v}_k\right\rangle \\
&= \left\langle \sum_{j=1}^{n} x_j \mathbf{v}_j, \sum_{k=1}^{n} y_k \mathbf{v}_k \right\rangle = \langle \mathbf{u}, \mathbf{v}\rangle .
\end{aligned}$$

(3) $\Rightarrow$ (1) Seja $\alpha = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$, uma base ortonormal de $\mathbb{R}^n$, e $a_{ij} \in \mathbb{R}$ escolhidos por satisfazer:

$$T(\mathbf{v}_j) = \sum_{i=1}^{n} a_{ij}\mathbf{v}_i \text{ e } T(\mathbf{v}_k) = \sum_{r=1}^{n} a_{rk}\mathbf{v}_r, \text{ com } j, k = 1, \ldots, n.$$

Isto é, os a_{ij} são os coeficientes da matriz $[T]_\alpha^\alpha$. Então,

$$\begin{aligned}
\delta_{ij} &= \langle \mathbf{v}_j, \mathbf{v}_k\rangle \\
&= \langle T(\mathbf{v}_j), T(\mathbf{v}_k)\rangle \\
&= \left\langle \sum_{i=1}^{n} a_{ij}\mathbf{v}_i, \sum_{r=1}^{n} a_{rk}\mathbf{v}_r \right\rangle \\
&= \sum_{i=1}^{n} a_{ij}a_{ik}.
\end{aligned}$$

Mas isso é equivalente a dizer que $[T]_\alpha^\alpha$ é ortogonal. $\square$

R9.3. a) Encontre a aplicação linear $S : \mathbb{R}^2 \to \mathbb{R}^2$, que é a reflexão em torno da reta $y = x$.

b) Obtenha a matriz de S na base canônica.

Solução: Lembrando que $S\mathbf{u} = 2P\mathbf{u} - \mathbf{u}$ é dada pela fórmula

$$S\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} \left(\frac{1-a^2}{1+a^2}\right) x + \left(\frac{2a}{1+a^2}\right) y \\ \left(\frac{2a}{1+a^2}\right) x + \left(\frac{a^2-1}{1+a^2}\right) y \end{bmatrix}.$$

Como a hipótese do exercício é $a = 1$, segue que $S\left(\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]\right) = \left[\begin{smallmatrix} y \\ x \end{smallmatrix}\right]$, e a matriz, em relação à base canônica, é

$$A = \left[S\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) \quad S\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)\right] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

□

R9.4. Prove o teorema 9.16: Se F é um subconjunto do $\mathbb{R}^n$, o qual está munido de um produto interno, então $F^\perp$ é um subespaço vetorial de $\mathbb{R}^n$.

Solução: Para garantir que $F^\perp$ é um subespaço vetorial de $\mathbb{R}^n$ é necessário verificar as seguintes três condições:

- $0 \in F^\perp$, uma vez que $\langle \mathbf{u}, 0 \rangle = 0$, para todo $\mathbf{u} \in F$;
- Se $\mathbf{v}, \mathbf{w} \in F^\perp$, então $\langle \mathbf{u}, \mathbf{v} + \alpha\mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \alpha \langle \mathbf{u}, \mathbf{w} \rangle = 0 + \alpha 0 = 0$ para todo $\alpha \in \mathbb{R}$ e, portanto, $\mathbf{v} + \alpha\mathbf{w} \in F^\perp$.

Isto é o suficiente para garantir que $F^\perp$ seja um subespaço vetorial. Veja que não exigimos nenhuma restrição ao subconjunto F. □

Exercícios propostos

P9.1. Encontre k, tal que os vetores

$$\mathbf{u} = \begin{bmatrix} 1 \\ -2 \\ k \\ 3 \end{bmatrix} \text{ e } \mathbf{v} = \begin{bmatrix} 3 \\ k \\ 7 \\ -6 \end{bmatrix}$$

de $\mathbb{R}^4$ sejam ortogonais.

P9.2. Encontre uma base ortogonal para os subespaços de $\mathbb{R}^3$, gerados pelos vetores:

a) $\left[\begin{smallmatrix} 1 \\ 1 \\ -1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ 0 \\ -1 \end{smallmatrix}\right]$ b) $\left[\begin{smallmatrix} 2 \\ 1 \\ 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ 3 \\ 1 \end{smallmatrix}\right]$.

P9.3. Em cada um dos casos abaixo, determine se o conjunto $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\} \subset \mathbb{R}^3$ é ortogonal, apenas ortogonal ou nenhum dos dois.

a) $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \mathbf{w} = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}.$

b) $\mathbf{u} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \mathbf{v} = \begin{bmatrix} -b \\ a \\ 0 \end{bmatrix}, \mathbf{w} = \begin{bmatrix} -ac \\ -bc \\ a^2+b^2 \end{bmatrix}.$

c) $\mathbf{u} = \frac{1}{7}\begin{bmatrix} 2 \\ 6 \\ 3 \end{bmatrix}, \mathbf{v} = \frac{1}{7}\begin{bmatrix} 3 \\ 2 \\ -6 \end{bmatrix}, \mathbf{w} = \frac{1}{7}\begin{bmatrix} 6 \\ -3 \\ 2 \end{bmatrix}.$

P9.4. Use a desigualdade de Cauchy-Schwarz no $\mathbb{R}^3$, para mostrar que, se $a > 0, b > 0$ e $c > 0$, então

$$(a+b+c)\left(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}\right) \geq 9.$$

P9.5. Sejam a, b e c números reais positivos, tais que $a + b + c = 1$. Use a desigualdade de Cauchy-Schwarz no $\mathbb{R}^3$, para mostrar que

$$\left(\frac{1}{a}-1\right)\left(\frac{1}{b}-1\right)\left(\frac{1}{c}-1\right) \geq 8.$$

P9.6. Encontre uma base ortogonal para o espaço solução de:

a) $\begin{cases} 2x + y + z = 0 \\ \quad\;\; y + z = 0 \end{cases}$ b) $x - y + z = 0.$

P9.7. a) Encontre a aplicação linear $S : \mathbb{R}^2 \to \mathbb{R}^2$, que é a reflexão em torno da reta $y = 2x$.

b) Obtenha a matriz de S na base canônica.

P9.8. Mostre a lei do paralelogramo: $|\mathbf{u}+\mathbf{v}|^2 + |\mathbf{u}-\mathbf{v}| = 2\,|\mathbf{u}|^2 + 2\,|\mathbf{v}|^2$, para quaisquer que sejam $\mathbf{u}, \mathbf{v} \in (\mathbb{R}^n, \langle\ ,\ \rangle)$.

P9.9. Sejam $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, prove que se $(\mathbf{u}+\mathbf{v}) \perp (\mathbf{u}-\mathbf{v})$, então $|\mathbf{u}| = |\mathbf{v}|$. Interprete o resultado geometricamente.

P9.10. Mostre que, se S é um conjunto ortogonal de vetores não nulos, S é linearmente independente.

P9.11. Seja $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ uma base ortogonal de $\mathbb{R}^n$. Então, dado qualquer $v \in \mathbb{R}^n$, verifique que as coordenadas de v com respeito a base são dadas por

$$\mathbf{v} = \frac{\langle \mathbf{v}, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle}\mathbf{u}_1 + \frac{\langle \mathbf{v}, \mathbf{u}_2 \rangle}{\langle \mathbf{u}_2, \mathbf{u}_2 \rangle}\mathbf{u}_2 + \cdots + \frac{\langle \mathbf{v}, \mathbf{u}_n \rangle}{\langle \mathbf{u}_n, \mathbf{u}_n \rangle}\mathbf{u}_n.$$

P9.12. Seja $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ uma base ortogonal de $\mathbb{R}^n$ e $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, dois vetores quaisquer, mostre que

$$\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{e}_1 \rangle \langle \mathbf{v}, \mathbf{e}_1 \rangle + \cdots + \langle \mathbf{u}, \mathbf{e}_n \rangle \langle \mathbf{v}, \mathbf{e}_n \rangle .$$

P9.13. Mostre que $(I-A)(I+A)^{-1}$ é uma matriz ortogonal, onde $A = \begin{bmatrix} 0 & 5 \\ -5 & 0 \end{bmatrix}$.

P9.14. Determine valores para x, y e z, tal que a matriz abaixo seja canônica

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ x & y & z \end{bmatrix}.$$

P9.15. Considere em $\mathbb{R}^3$ a função dada por $\langle \mathbf{u}, \mathbf{v} \rangle = 8x_1x_2 - 3x_2y_1 - 3x_1y_2 + 2y_1y_2 + 2z_1z_2$, onde $\mathbf{u} = \left[\begin{smallmatrix} x_1 \\ y_1 \\ z_1 \end{smallmatrix}\right]$ e $\mathbf{v} = \left[\begin{smallmatrix} x_2 \\ y_2 \\ z_2 \end{smallmatrix}\right]$.

(a) Verifique que esaa função é um produto interno.

(b) Usando a norma que é a consequência deste produto interno, encontre o vetor do plano $2x + 3y - 6z = 0$ que está mais próximo do vetor $\mathbf{w} = \left[\begin{smallmatrix} 1 \\ 1 \\ -3 \end{smallmatrix}\right]$.

(c) Calcule a distância, usando a função distância que provem deste produto interno, de $\mathbf{w}$ até o plano $2x + 3y - 6z = 0$.

P9.16. Considere o $\mathbb{R}^2$ com o produto interno dado por $\langle \mathbf{u}, \mathbf{v} \rangle = x_1y_1 + 2x_2y_2 - x_1y_2 - x_2y_1$, onde $\mathbf{u} = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]$ e $\mathbf{v} = \left[\begin{smallmatrix} y_1 \\ y_2 \end{smallmatrix}\right]$.

(a) Determine m, de tal forma que os vetores $\left[\begin{smallmatrix} 1+m \\ 2 \end{smallmatrix}\right]$ e $\left[\begin{smallmatrix} 3 \\ m-1 \end{smallmatrix}\right]$ sejam ortogonais.

(b) Determine todos os vetores de $\mathbb{R}^2$ ortogonais a $\left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$.

(c) Determine todos os vetores $\left[\begin{smallmatrix} m \\ m-1 \end{smallmatrix}\right]$ que têm norma 1.

P9.17. Mostre que uma transformação ortogonal do plano no plano deixa invariante a distância entre dois pontos, isto é, dados $\mathbf{u}$ e $\mathbf{v}$, dois vetores quaisquer,

$$\|T\mathbf{u} - T\mathbf{v}\| = \|\mathbf{u} - \mathbf{v}\|.$$

P9.18. Mostre que, se T é uma transformação ortogonal do plano no plano, sua matriz em relação à base canônica só pode ser da forma:

$$A = \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix} \text{ (Rotações)}$$

ou da forma

$$B = \begin{bmatrix} \cos\theta & \operatorname{sen}\theta \\ \operatorname{sen}\theta & -\cos\theta \end{bmatrix} \text{ (Reflexões).}$$

Capítulo 10

Cônicas, Matrizes Simétricas e Formas Quadráticas

Nesse capítulo vamos exibir dois métodos de mudança de coordenadas. No primeiro método mostraremos um resultado mais forte, o qual nos fala que: para toda equação do 2º grau, existe um sistema de coordenadas ortonormal no qual a equação se escreve como uma das equações padrões das cônicas. Para demonstrar esse resultado, introduziremos um tipo especial de operador linear, conhecido como operador autoadjunto, e mostraremos que todo operador desse tipo possui uma base ortonormal de autovetores.

O segundo método dar-nos-á um algoritmo que permite reduzir qualquer forma quadrática a uma forma diagonal, consistindo em uma técnica útil para cálculos em espaços de dimensão superior. Além disso, essa técnica é usada para classificar as formas bilineares simétricas. Ela tem o inconveniente, porém, de gerar um sistema de coordenadas final que não é ortogonal.

Ao final do capítulo aplicaremos tais técnicas para classificar as quadráticas em duas variáveis. Deve ficar claro que esse processo pode ser estendido para equações de segundo grau em mais de duas variáveis.

Além disso, neste capítulo, nós iremos introduzir três conceitos que estão relacionados: matrizes simétricas, formas bilineares simétricas e formas quadráticas. No texto, apresentaremos as conexões entre eles, mas não nos aprofundaremos no estudo das formas quadráticas e na classificação das formas bilineares. Gostaríamos apenas de chamar a atenção para o fato de que esses dois conceitos oferecem inúmeras aplicações em questões de otimização e de tratamento de dados.

10.1 Cônicas

As cônicas, ou secções cônicas, foram estudadas pelos gregos e desenvolvidas a partir de suas propriedades geométricas. Pouco se sabe a respeito dos precursores do estudo das cônicas. Restaram apenas os trabalhos de Apollonius de Perga (260-170 A. C.). Nesses trabalhos, não só foram compilados os resultados conhecidos na época como também apresentada de forma sistemática a dedução das propriedades das cônicas.

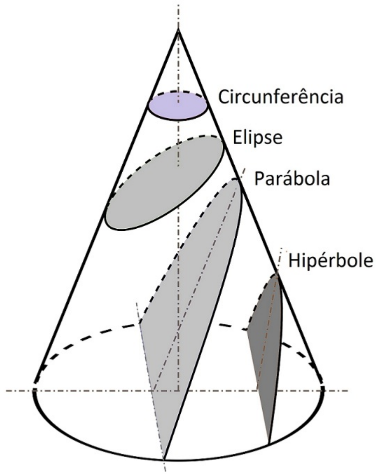


Figura 10.1: Seções de um Cone

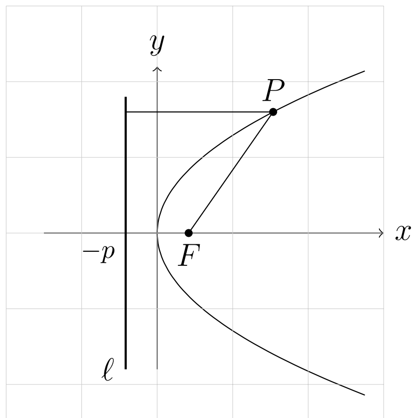
Abaixo daremos a definição das cônicas proposta por Apollonius.

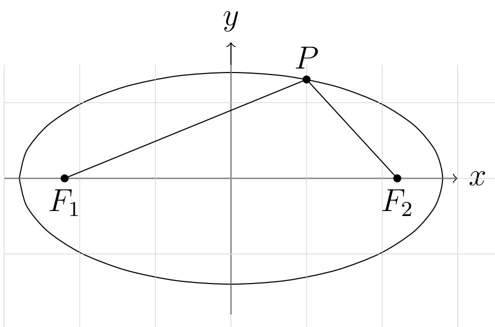
Definição 10.1

A **parábola** é obtida por considerar os pontos P do plano que satisfazem

$$\text{dist}(F, P) = \text{dist}(P, \ell),$$

onde ℓ é a **reta diretriz**, e sua equação é $x = -p$.





A **elipse** é obtida por considerar os pontos P do plano que satisfazem

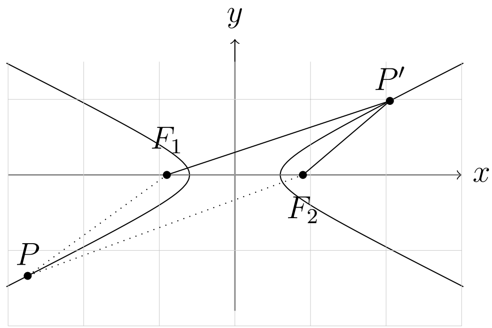
$$\text{dist}(P, F_1) + \text{dist}(P, F_2) = 2a. \tag{10.1}$$

A **hipérbole** é obtida por considerar os pontos P, P' do plano que satisfazem

$$\text{dist}(P', F_1) - \text{dist}(P', F_2) = 2a. \quad (10.2)$$

ou

$$\text{dist}(P, F_1) - \text{dist}(P, F_2) = -2a. \quad (10.3)$$



É possível, a partir das definições acima, obter as seguintes equações padrões:

$$y^2 = 4px, \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \text{ ou } \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1,$$

as quais descrevem uma parábola, uma elipse, ou uma hipérbole, respectivamente.

Para exemplificar como é feita a passagem da definição acima até a equação, vamos tratar de forma completa somente o caso da parábola, que, das três, é a mais simples de deduzir.

10.1.1 Equação da Parábola

Considere $a > 0$. Defina $p = a = c$ e introduza um sistema de coordenada xy tal que o eixo do x passe por cima dos focos e esse tenha coordenadas $F = \left[\begin{smallmatrix} p \\ 0 \end{smallmatrix}\right]$, e o eixo do y seja paralela à reta diretriz ℓ e satisfaça a equação $x = -p$. Da definição da parábola

$$\text{dist}(F, P) = \text{dist}(P, \ell),$$

se $P = \left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$, segue que $\text{dist}(F, P) = \sqrt{(x-p)^2 + y^2}$ e $\text{dist}(P, \ell) = |x + p|$. A equação fica

$$\sqrt{(x-p)^2 + y^2} = |x + p|.$$

Elevando ao quadrado ambos o membros dessa equação, e simplificando, obtemos:

$$y^2 = 4px.$$

10.1.2 Equação da Elipse

Considere $a > b > 0$. Defina $c = \sqrt{a^2 - b^2}$ que é conhecida como a **distância focal da elipse**. Vamos introduzir um sistema de coordenadas xy, de tal forma que os focos da elipse fiquem sobre o eixo x e tenham coordenadas $F_1 = \left[\begin{smallmatrix} -c \\ 0 \end{smallmatrix}\right]$ e $F_2 = \left[\begin{smallmatrix} c \\ 0 \end{smallmatrix}\right]$, denominados focos à esquerda e à direita, respectivamente. Escolha o eixo do y de tal forma que ele seja a mediatriz do segmento F_1F_2. Com esse sistema de coordenadas, a equação 10.1 torna-se

$$\sqrt{(x+c)^2 + y^2} + \sqrt{(x-c)^2 + y^2} = 2a.$$

Assim, passando o segundo radical para o outro lado da igualdade, elevando ao quadrado, simplificando, e usando a relação $c^2 = a^2 - b^2$, obtemos a equação padrão da elipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

10.1.3 Equação da Hipérbole

Considere $a > b > 0$. Defina $c = \sqrt{a^2 + b^2}$ que é conhecido como a **distância focal da Hipérbole**. Aqui também introduza um sistema de coordenada xy tal que o eixo do x passe por cima dos focos que tenham coordenadas $F_1 = \left[\begin{smallmatrix} -c \\ 0 \end{smallmatrix}\right]$ e $F_2 = \left[\begin{smallmatrix} c \\ 0 \end{smallmatrix}\right]$, denominados focos à esquerda e à direita. Escolha o eixo de y de modo que ele seja a mediatriz do segmento F_1F_2. Com esse sistema, as equações 10.2 e 10.3 tornam-se, respectivamente,

$$\sqrt{(x+c)^2 + y^2} - \sqrt{(x-c)^2 + y^2} = 2a \text{ e } \sqrt{(x+c)^2 + y^2} - \sqrt{(x-c)^2 + y^2} = -2a.$$

Escolhendo qualquer uma dessas equações, procedendo de forma idêntica ao que foi feito com a equação da elipse, e usando que $c^2 = a^2 + b^2$, obtemos a equação padrão da hipérbole

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1.$$

10.1.4 Excentricidade e Propriedades das Cônicas

A seguir iremos definir alguns conceitos importantes a respeito das cônicas. Diremos que as diretrizes da elipse e da hipérbole são as retas:

$$y = -\frac{a}{e} \text{ e } x = \frac{a}{e}, \text{ respectivamente. A circunferência não tem diretriz.}$$

Observação 10.2
Se tivermos na elipse ou na hipérbole $b > a > 0$, a distância focal será, respectivamente, $c = \sqrt{b^2 - a^2}$, $c = \sqrt{a^2 + b^2}$. Devemos escolher o sistema de coordenadas de maneira que os focos tenham coordenadas $F_1 = \left[\begin{smallmatrix} 0 \\ -c \end{smallmatrix}\right]$ e $F_2 = \left[\begin{smallmatrix} 0 \\ c \end{smallmatrix}\right]$ e, nesse caso, as diretrizes serão $y = \pm\frac{b}{e}$.

As cônicas possuem um invariante que nos permite defini-las mais sinteticamente e nos oferece uma outra visão sobre a natureza delas. Para esclarecer melhor isso, introduziremos a noção de **excentricidade** e, que é, por definição, $e = \frac{c}{a}$.

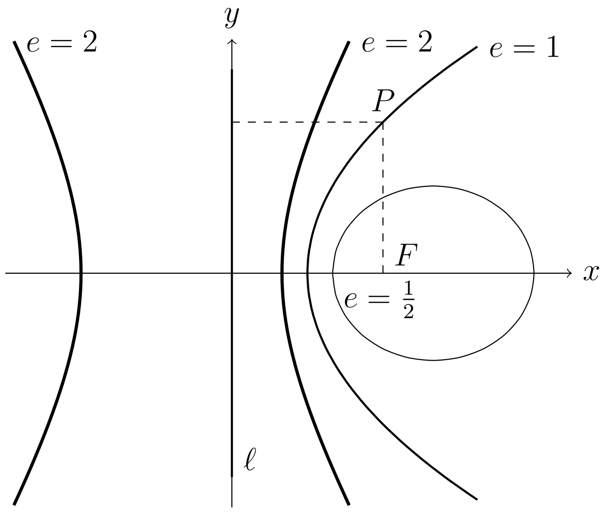
Segue da definição de excentricidade que, se a cônica for uma elipse, então, $0 \leq e < 1$ (na circunferência, e = 0); e se for uma hipérbole, então $e > 1$, e se for uma parábola, então, $e = 1$.

Proposição 10.3
Fixe um ponto F (Foco), uma linha ℓ (a diretriz), que não contenha F, e escolha um número Real positivo e (a excentricidade). Considere que os pontos P satisfaçam

$$\operatorname{dist}(P, F) = e \operatorname{dist}(P, \ell).$$

Mostre que: a) se $0 < e < 1$, obtemos uma elipse; b) se $e = 1$, obtemos uma parábola; c) se $e > 1$, obtemos uma hipérbole.

Demonstração: veja o exercício R10.1 e figura 10.1.4 que ilustra a propriedade. □



Nos exercícios, exibiremos outras propriedades interessantes das cônicas. No intuito de mostrar que toda equação do 2ª grau pode ser vista como uma cônica em um sistema de coordenadas ortonormal, vamos apresentar uma família de operadores lineares, conhecida como família de operadores autoadjuntos.

10.2 Operadores Autoadjuntos

Seja $\langle , \rangle$ um produto interno em $\mathbb{R}^n$. Diremos que um operador linear $T : \mathbb{R}^n \to \mathbb{R}^n$ será um operador autoadjunto se a matriz $[T]^\alpha_\alpha$, com respeito à uma base ortonormal α, for simétrica.

Suponha que $T : \mathbb{R}^2 \to \mathbb{R}^2$ seja dado por $\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] \mapsto \left[\begin{smallmatrix} 3x+4y \\ 4x-7y \end{smallmatrix}\right]$. Considerando $\mathbb{R}^2$ com o produto interno usual e α a base canônica do $\mathbb{R}^2$, temos que a matriz

$$[T]^\alpha_\alpha = \begin{bmatrix} 3 & 4 \\ 4 & -7 \end{bmatrix}$$

é simétrica e, portanto, esse operador é um operador autoadjunto.

Teorema 10.4
Sejam $\langle \, , \, \rangle$ um produto interno em $\mathbb{R}^n$ e $T : \mathbb{R}^n \to \mathbb{R}^n$ um operador linear. Então, T é um operador autoadjunto se, e somente se,

$$\langle T\mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, T\mathbf{v} \rangle \,, \text{ para todo } \mathbf{u}, \mathbf{v} \in \mathbb{R}^n.$$

Demonstração: veja o exercício R10.2 □

Vamos verificar tal propriedade com a transformação $T : \mathbb{R}^2 \to \mathbb{R}^2$ definida acima e com o produto interno usual do $\mathbb{R}^2$. Considere $\mathbf{u} = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]$ e $\mathbf{v} = \left[\begin{smallmatrix} y_1 \\ y_2 \end{smallmatrix}\right]$ então,

$$\begin{aligned} \langle T\mathbf{u}, \mathbf{v} \rangle &= \left\langle \left[\begin{smallmatrix} 3x_1+4x_2 \\ 4x_1-7x_2 \end{smallmatrix}\right], \left[\begin{smallmatrix} y_1 \\ y_2 \end{smallmatrix}\right] \right\rangle \\ &= 3x_1y_1 + 4x_2y_1 + 4x_1y_2 - 7x_2y_2 \\ &= 3x_1y_1 + 4x_1y_2 + 4x_2y_1 - 7x_2y_2 \\ &= x_1(3y_1 + 4y_2) + x_2(4y_1 - 7y_2) \\ &= \left\langle \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right], \left[\begin{smallmatrix} 3y_1+4y_2 \\ 4y_1-7y_2 \end{smallmatrix}\right] \right\rangle = \langle \mathbf{u}, T\mathbf{v} \rangle \,. \end{aligned}$$

Corolário 10.5
Suponha que $T : \mathbb{R}^n \to \mathbb{R}^n$ seja um operador autoadjunto; então a matriz de T será simétrica com respeito a qualquer base ortonormal.

Proposição 10.6
Seja $\langle\, , \,\rangle$ um produto interno em $\mathbb{R}^n$. Se $T : \mathbb{R}^n \to \mathbb{R}^n$ é um operador autoadjunto então:
a) autovetores correspondentes a autovalores distintos são ortogonais;
b) T sempre possui pelo menos um autovalor.

Demonstração: veja o exercício R10.3 □

Lema 10.7
Seja $T : \mathbb{R}^n \to \mathbb{R}^n$ um operador autoadjunto. Se o subespaço $W \subset \mathbb{R}^n$ é T-invariante, então seu complemento ortogonal $W^\perp$ também o é.

Demonstração: considere $\mathbf{u} \in W$ e $\mathbf{v} \in W^\perp$, então $T\mathbf{u} \in W$ e

$$\langle \mathbf{u}, T\mathbf{v}\rangle = \langle T\mathbf{u}, \mathbf{v}\rangle = 0.$$

Mas isso implica que $T\mathbf{v} \in W^\perp$, logo $W^\perp$ também é T-invariante. □

Teorema 10.8 (Teorema Espectral)
Seja $\langle\, , \,\rangle$ um produto interno em $\mathbb{R}^n$. Se $T : \mathbb{R}^n \to \mathbb{R}^n$ é um operador autoadjunto, então existe uma base ortogonal de $\mathbb{R}^n$ composta por autovetores de T.

Demonstração: veja o exercício R10.5. □

Exemplo 10.9
Considere $\mathbb{R}^2$ com o produto interno usual e $T\left(\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]\right) = \left[\begin{smallmatrix} x+3y \\ 3x-7y \end{smallmatrix}\right]$. Seja $\mathcal{C}$ a base canônica do $\mathbb{R}^2$, chame

$$A = [T]_{\mathcal{C}}^{\mathcal{C}} = \begin{bmatrix} 1 & 3 \\ 3 & -7 \end{bmatrix}. \text{ Segue que } \Delta_A(x) = x^2 + 6x - 16 = (x-2)(x+8).$$

Como a matriz A é simétrica, segue que T é autoadjunto e, portanto, pelo teorema 10.8 (Teorema Espectral) este operador admite uma base de autovetores ortonormal. Vamos determiná-la.
(i) Por subtrair $\lambda = 2$ da diagonal da matriz A, obtemos

$$\begin{bmatrix} -1 & 3 \\ 3 & -9 \end{bmatrix} \left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] = \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right] \Leftrightarrow -x + 3y = 0.$$

E, portanto, um autovalor associado a $\lambda = 2$ é $\left[\begin{smallmatrix} 3y \\ y \end{smallmatrix}\right] = y\left[\begin{smallmatrix} 3 \\ 1 \end{smallmatrix}\right]$.
(ii) Ao invés de seguir o procedimento padrão, vamos usar a informação que o Teorema Espectral nos fornece. Sabemos que os autovetores associado ao autovalor $\lambda = 2$ são ortogonais à $\left[\begin{smallmatrix} 3 \\ 1 \end{smallmatrix}\right]$, e, como estamos em $\mathbb{R}^2$, um autovetor associado a $\lambda = -8$ deve ser $\left[\begin{smallmatrix} -1 \\ 3 \end{smallmatrix}\right]$. De fato, veja que

$$\begin{bmatrix} 1 & 3 \\ 3 & -7 \end{bmatrix} \left[\begin{smallmatrix} -1 \\ 3 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 8 \\ -24 \end{smallmatrix}\right] = -8\left[\begin{smallmatrix} -1 \\ 3 \end{smallmatrix}\right].$$

Normalizando esses vetores, obtemos uma base ortonormal $\beta = \left\{ \begin{bmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \end{bmatrix}, \begin{bmatrix} -1/\sqrt{10} \\ 3/\sqrt{10} \end{bmatrix} \right\}$. A matriz mudança de coordenadas

$$P = [I]_\alpha^\beta = \begin{bmatrix} \frac{3}{\sqrt{10}} & \frac{-1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix}.$$

Como α e β são bases ortonormais, temos que P é ortogonal, isto é, $P^t = P^{-1}$. com base nisso, teremos

$$D = \begin{bmatrix} 2 & 0 \\ 0 & -8 \end{bmatrix} = P^t A P.$$

10.3 Formas Bilineares Simétricas e Formas Quadráticas

Diremos que uma aplicação $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ será bilinear se ela satisfizer as seguintes propriedades para $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ e $\alpha \in \mathbb{R}$ quaisquer,

$$f(\mathbf{u} + \alpha\mathbf{w}, \mathbf{v}) = f(\mathbf{u}, \mathbf{v}) + \alpha f(\mathbf{w}, \mathbf{v}) \text{ e } f(\mathbf{u}, \mathbf{v} + \alpha\mathbf{w}) = f(\mathbf{u}, \mathbf{v}) + \alpha f(\mathbf{u}, \mathbf{w}).$$

Além disso, diremos que ela é simétrica se

$$f(\mathbf{u}, \mathbf{v}) = f(\mathbf{v}, \mathbf{u}) \text{ para quaisquer que sejam } \mathbf{u}, \mathbf{v} \in \mathbb{R}^n.$$

Como exemplos de formas bilineares simétrica, podemos citar os produtos internos sobre $\mathbb{R}^n$.

Às formas bilineares simétricas, podemos associar, para cada base, uma matriz simétrica A da seguinte forma: fixe uma base $\alpha = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ de $\mathbb{R}^n$, tome $\mathbf{u}$ e $\mathbf{v}$ vetores de $\mathbb{R}^n$. Calculando as suas coordenadas, temos $\mathbf{u} = \sum_{i=1}^n x_i \mathbf{v}_i$, $\mathbf{v} = \sum_{j=1}^n y_j \mathbf{v}_j$ então,

$$\begin{aligned} f(\mathbf{u}, \mathbf{v}) &= f\left(\sum_{i=1}^n x_i \mathbf{v}_i, \sum_{j=1}^n y_j \mathbf{v}_j\right) \\ &= \sum_{i=1}^n x_i f\left(\mathbf{v}_i, \sum_{j=1}^n y_j \mathbf{v}_j\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n x_i y_j f(\mathbf{v}_i, \mathbf{v}_j). \end{aligned}$$

Portanto, para calcular $f(\mathbf{u}, \mathbf{v})$, basta saber os valores $f(\mathbf{v}_i, \mathbf{v}_j)$, $1 \leq i, j \leq n$. Se criarmos a matriz (simétrica) $A = [f]_\alpha = [f(\mathbf{v}_i, \mathbf{v}_j)]$, isto é, a matriz que na posição ij tem o valor $f(\mathbf{v}_i, \mathbf{v}_j)$, então,

$$f(\mathbf{u}, \mathbf{v}) = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \begin{bmatrix} f(\mathbf{v}_1, \mathbf{v}_1) & f(\mathbf{v}_1, \mathbf{v}_2) & \cdots & f(\mathbf{v}_1, \mathbf{v}_n) \\ f(\mathbf{v}_2, \mathbf{v}_1) & f(\mathbf{v}_2, \mathbf{v}_2) & \cdots & f(\mathbf{v}_2, \mathbf{v}_n) \\ \vdots & \vdots & \ddots & \vdots \\ f(\mathbf{v}_n, \mathbf{v}_1) & f(\mathbf{v}_n, \mathbf{v}_2) & \cdots & f(\mathbf{v}_n, \mathbf{v}_n) \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

Chamamos a matriz $A = [f]_\alpha = [f(\mathbf{v}_i, \mathbf{v}_j)]$ de **matriz associada à forma bilinear simétrica na base** α.

Exemplo 10.10
Considere a forma bilinear f do $\mathbb{R}^3$ definida por

$$f\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}\right) = 2x_1y_1 - 4x_2y_1 - 8x_3y_1 - 4x_1y_2 + x_2y_2 + 7x_3y_2 - 8x_1y_3 + 7x_2y_3 + 5x_3y_3.$$

Seja $\mathcal{C} = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ a base canônica do $\mathbb{R}^3$. Calculando $f(\mathbf{e}_1, \mathbf{e}_1) = 2$, $f(\mathbf{e}_1, \mathbf{e}_2) = -4$ e todas as outras possibilidades, obtemos a matriz

$$[f]_{\mathcal{C}} = \begin{bmatrix} 2 & -4 & -8 \\ -4 & 1 & 7 \\ -8 & 7 & 5 \end{bmatrix}.$$

Definição 10.11
Dizemos que uma aplicação $Q : \mathbb{R}^n \to \mathbb{R}$ será uma forma quadrática, se $Q(\mathbf{u}) = f(\mathbf{u}, \mathbf{u})$ para alguma forma bilinear simétrica f de $\mathbb{R}^n$.

Há uma maneira de, considerando uma forma quadrática Q, obtermos a forma bilinear f necessária para a definição 10.11. Considere a aplicação $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ dada por

$$f(\mathbf{u}, \mathbf{v}) = \frac{1}{2}\left[Q(\mathbf{u} + \mathbf{v}) - Q(\mathbf{u}) - Q(\mathbf{v})\right] \text{ (Fórmula de Polarização).}$$

Para maiores detalhes, veja o exercício R10.6.

Portanto, dado uma forma quadrática $Q : \mathbb{R}^n \to \mathbb{R}$, podemos determinar uma forma bilinear simétrica f, através da fórmula de polarização. Além disso, se fixarmos uma base α de $\mathbb{R}^n$ podemos associar uma matriz $A = [f]_\alpha$ à forma bilinear f. Por isso, também, chamamos a matriz A de **matriz da forma quadrática** Q **com respeito à base** α.

É possível fazer o caminho inverso. Se tivermos uma matriz $A = [a_{ij}]_{n \times n}$, simétrica, poderemos determinar uma forma quadrática fazendo $Q(\mathbf{u}) = \mathbf{u}^t A \mathbf{u}$.

Observação 10.12
Seja f uma forma bilinear representada pela matriz simétrica $A = [a_{ij}]$. Considere $\mathbf{u} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$. Então a forma quadrática Q pode ser representada por

$$\begin{aligned} Q(\mathbf{u}) = f(\mathbf{u}, \mathbf{u}) &= \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= \sum_{i,j} a_{ij} x_i x_j = \sum_i a_{ii} x_i^2 + \sum_{i<j} a_{ij} x_i x_j. \end{aligned}$$

Segue que Q será uma forma quadrática se, e somente se, $Q(\mathbf{u})$ for um polinômio homogêneo de grau 2 nas entradas do vetor $\mathbf{u}$.

Pela observação 10.12, dada a aplicação $\mathbb{Q} : \mathbb{R}^2 \to \mathbb{R}$, definida por

$$Q\left(\left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]\right) = ax_1^2 + bx_1x_2 + bx_2^2, \text{ onde } a,\, b,\, c \in \mathbb{R},$$

temos que esta é uma forma quadrática, pois consiste de um polinômio homogêneo de grau 2. Além disso, é fácil verificar que

$$Q\left(\left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]\right) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Essa observação se aplica a qualquer $n \geq 3$, veja o exemplo 10.13.

Exemplo 10.13
Encontre a matriz simétrica A, associada à forma quadrática

$$Q\left(\left[\begin{smallmatrix} x \\ y \\ z \end{smallmatrix}\right]\right) = 2x^2 - 8xy + y^2 - 16xz + 14yz + 5z^2.$$

Veja que os coeficientes que aparecem na diagonal são os mesmos que acompanham os termos x^2, y^2 e z^2. Os coeficientes que acompanham os outros termos devem aparecer, divididos por 2, fora da diagonal principal. Por exemplo: -8 que é o coeficiente que acompanha xy deve aparecer, dividido por 2, na linha 1 coluna 2 e na linha 2 coluna 1. Então,

$$Q\left(\left[\begin{smallmatrix} x \\ y \\ z \end{smallmatrix}\right]\right) = \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} 2 & -4 & -8 \\ -4 & 1 & 7 \\ -8 & 7 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

A matriz acima é matriz de Q com respeito à base canônica do $\mathbb{R}^3$ (veja exemplo 10.10).

10.3.1 Mudança de Coordenada e Formas Bilineares Simétricas

Suponha que $\mathbf{x}, \mathbf{y}$ são as coordenadas de dois vetores do $\mathbb{R}^n$ em relação à base $\alpha = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$. Digamos que $\mathbf{x}'$ e $\mathbf{y}'$ representam os mesmos os vetores em relação a outra base $\beta = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. Dessa forma, existe uma matriz invertível $P = [I]_\alpha^\beta$, que satisfaz

$$\mathbf{x} = P\mathbf{x}' \text{ e, de outra forma, } \mathbf{y} = P\mathbf{y}'.$$

Se $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ é uma forma bilinear simétrica e $A = [f(\mathbf{u}_i, \mathbf{u}_j)]$, então:

$$f(\mathbf{x}, \mathbf{y}) = \mathbf{x}^t A\mathbf{y} = (P\mathbf{x}')^t AP\mathbf{y}' = \mathbf{x}'^t P^t AP\mathbf{y}' = f(\mathbf{x}', \mathbf{y}').$$

Segue que a matriz de f, com respeito a base β, isto é, $B = [f(\mathbf{v}_i, \mathbf{v}_j)]$, deve satisfazer $B = P^tAP$.

Definição 10.14
Diremos que uma matriz B será **congruente** a uma matriz A (e escrevemos $B \cong A$), se existir, uma matriz invertível P tal que $B = P^tAP$.

Observação 10.15
Não confunda o conceito de congruência (definição 10.14) entre matrizes com o de semelhança (definição 7.11) entre matrizes. Apesar disso, se a matriz P for ortogonal, os dois conceitos irão coincidir.

De acordo com a definição acima, podemos dizer que matrizes associadas a uma mesma forma quadrática, em relação a bases diferentes, serão sempre congruentes, e, reciprocamente, matrizes congruentes estarão associadas a uma mesma forma quadrática, só que relacionadas a outro sistema de coordenadas.

Exemplo 10.16
Considere $\mathbf{u} = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right] \in \mathbb{R}^2$ e a forma quadrática $q : \mathbb{R}^2 \to \mathbb{R}$, definida por $q(\mathbf{u}) = x_1^2 + 6x_1x_2 - 7x_2^2$. Esta pode ser reescrita, em termos da matriz associada, da seguinte forma:

$$q(\left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 3 & -7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Se escolhemos outra base, digamos $\beta = \left\{ \begin{bmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \end{bmatrix}, \begin{bmatrix} -1/\sqrt{10} \\ 3/\sqrt{10} \end{bmatrix} \right\}$ (veja o exemplo 10.9). Sabemos que a base inicial era a canônica, portanto a matriz de mudança de coordenadas é

$$P = [I]_\alpha^\beta = \begin{bmatrix} \frac{3}{\sqrt{10}} & \frac{-1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix}.$$

Se $[\mathbf{u}]_\beta = \left[\begin{smallmatrix} y_1 \\ y_2 \end{smallmatrix}\right]$ são as coordenadas do vetor $\mathbf{u}$ com respeito à base β, então:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{3}{\sqrt{10}} & \frac{-1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \text{ e } \begin{bmatrix} x_1 & x_2 \end{bmatrix} = \begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{bmatrix} \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{-1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix}.$$

Dessa forma, temos que

$$\begin{aligned} q(\mathbf{u}) = q(\left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]) &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 3 & -7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{bmatrix} \frac{3}{\sqrt{10}} & \frac{-1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix}^t \begin{bmatrix} 1 & 3 \\ 3 & -7 \end{bmatrix} \begin{bmatrix} \frac{3}{\sqrt{10}} & \frac{-1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \\ &= \begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & -8 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 2y_1^2 - 8y_2^2 = q(\left[\begin{smallmatrix} y_1 \\ y_2 \end{smallmatrix}\right]). \end{aligned}$$

Quem já estudou equações do 2º grau ($Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$) sabe que a principal dificuldade é retirar o fator xy, mas as contas acima nos sugerem uma técnica para resolvermos esse problema.

10.4 Algoritmo de Diagonalização

Apesar de já conhecermos um procedimento para diagonalizar matrizes simétricas, vamos apresentar um algoritmo que nos permite diagonalizar uma matriz simétrica sem precisar passar pelo processo de calcular o polinômio característico, os autovalores e autovetores. Esse algoritmo exige apenas o conhecimento de operações elementares. Contra esse procedimento, pesa o fato de que a base na qual a matriz simétrica se torna diagonal nem sempre é ortonormal.

Vamos exibir um algoritmo que permite obter uma matriz invertível P que diagonaliza a matriz A. Existe uma sequência de operações elementares aplicado às linhas e uma forma de aplicar, essas operações, às colunas, que transforma a matriz A na matriz diagonal D, além disso, $D = P^t AP$.

Antes de iniciar a leitura do procedimento, veja o exemplo ilustrativo a seguir.

Observação 10.17
Considere a matriz simétrica

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

Recordemos que se quisermos fazer uma operação elementar entre as linhas, por exemplo, $\ell_2 \to \ell_2 - 2\ell_1$, basta realizarmos a mesma operação na matriz identidade, obtendo uma matriz que chamamos de matriz elementar E. Então, se fizermos EA, obteremos a mesma matriz que havíamos obtido quando fizemos a operação elementar em A. Além disso, se tomarmos E^t e fizermos AE^t, obteremos a mesma matriz A que sofreu a operação $\mathbf{c}_2 \to \mathbf{c}_2 - 2\mathbf{c}_1$ nas colunas.

Veja o exemplo:

$$EA = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \text{ e } AE^t = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}^t = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}.$$

Tendo ele como parâmetro, vamos descrever o algoritmo:

Algoritmo

Passo 0 Dada a matriz simétrica $A = [a_{ij}]$ de ordem n.

Passo 1 Forme uma matriz $n \times 2n$ em blocos $M = [A_1 : I]$, em que $A_1 = A$ é a metade à esquerda de M e a metade à direita de M é a matriz identidade.

Passo 2 Ao examinar o termo a_{11} de M há três possibilidades:

Caso I Se $a_{11} \neq 0$, então faça as operações $\ell_i \to \ell_i - \frac{a_{i1}}{a_{11}}\ell_1$ com $i = 2, 3, \ldots, n$ e, depois, faça o mesmo com as colunas, isto é, $\mathbf{c}_i \to \mathbf{c}_i - \frac{a_{i1}}{a_{11}}\mathbf{c}_1$ com $i = 2, 3, \ldots, n$. Essas operações irão reduzir a matriz M à forma

$$M \approx \begin{bmatrix} a_{11} & 0 & * & * \\ 0 & A_2 & * & * \end{bmatrix}.$$

Caso II - Se $a_{11} = 0$, mas $a_{kk} \neq 0$, para algum $k > 1$. Nesse caso, faça $\ell_1 \leftrightarrow \ell_k$ e, também, $\mathbf{c}_1 \leftrightarrow \mathbf{c}_k$.
Essas operações reduzem a matriz ao Caso I.
Caso III - Todas as entradas diagonais $a_{ii} = 0$, mas algum $a_{ij} \neq 0$. Nesse caso, faça $\ell_i \to \ell_j + \ell_i$ e o mesmo na coluna $\mathbf{c}_i \to \mathbf{c}_j + \mathbf{c}_i$ (Essas operações levam $2a_{ij}$ para a i-ésima entrada diagonal). Assim M volta à situação do Caso II.

Passo 3 Repita o passo 2 com cada nova matriz A_k (suprimindo as primeiras linhas e colunas da matriz precedente) até que A esteja diagonalizada. Então, M torna-se $M' = [D : Q]$, onde D é diagonal.

Passo 4 Tome $P = Q^t$. Então, $D = P^tAP$.

Observação 10.18
No passo 2, as operações com as linhas modificam os dois lados de M, mas as operações com as colunas modificam apenas o lado esquerdo de M.

Exemplo 10.19
Seja $A = \begin{bmatrix} 1 & 2 & -3 \\ 2 & 5 & -4 \\ -3 & -4 & 8 \end{bmatrix}$. Utilizando, nas operações, o algoritmo anterior, vamos determinar uma matriz não singular P tal que $D = P^tAP$ seja diagonal.

Iniciamos contruindo a matriz $M = [A : I]$

$$M = [A : I] = \begin{bmatrix} 1 & 2 & -3 & 1 & 0 & 0 \\ 2 & 5 & -4 & 0 & 1 & 0 \\ -3 & -4 & 8 & 0 & 0 & 1 \end{bmatrix}.$$

Aplicando as operações $\ell_2 \to \ell_2 - 2\ell_1$ e $\ell_3 \to \ell_3 + 3\ell_1$ nas linhas de M, e depois, aplicando as operações correspondentes $\mathbf{c}_2 \to \mathbf{c}_2 - 2\mathbf{c}_1$ e $\mathbf{c}_3 \to \mathbf{c}_3 + 3\mathbf{c}_1$ nas colunas, obtemos

$$\begin{bmatrix} 1 & 2 & -3 & 1 & 0 & 0 \\ 0 & 1 & 2 & -2 & 1 & 0 \\ 0 & 2 & -1 & 3 & 0 & 1 \end{bmatrix} \text{ e depois, } \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & -2 & 1 & 0 \\ 0 & 2 & -1 & 3 & 0 & 1 \end{bmatrix}.$$

Em seguida, faremos $\ell_3 \to \ell_3 - 2\ell_2$ e a operação correspondente $\mathbf{c}_3 \to \mathbf{c}_3 - 2\mathbf{c}_2$ obtendo, assim,

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & -2 & 1 & 0 \\ 0 & 0 & -5 & 7 & -2 & 1 \end{bmatrix} \text{ e depois, } \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -2 & 1 & 0 \\ 0 & 0 & -5 & 7 & -2 & 1 \end{bmatrix}.$$

Dessa forma, A foi diagonalizada. Escrevemos

$$P = \begin{bmatrix} 1 & -2 & 7 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix} \text{ e, portanto, } D = P^tAP = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -5 \end{bmatrix}.$$

Observe que P é a transposta da metade direita da matriz final.

Por fim, vamos justificar o algoritmo. Considere a matriz em blocos $M = [A : I]$. O algoritmo aplica uma sequência de operações elementares nas linhas, seguida de uma sequência de operações nas colunas do lado esquerdo de M, que é a matriz A. Isso equivale a pré-multiplicar A por uma sequência de matrizes elementares, digamos $E_1, E_2, \dots, E_k$ e, depois, a multiplicar A pelas transpostas das $E_{i's}$. Assim, quando o algoritmo terminar, a matriz diagonal D, à esquerda de M, será igual a

$$D = E_k \cdots E_2 E_1 A E_1^t E_2^t \cdots E_k^t, \text{ com } Q = E_k \cdots E_2 E_1.$$

Por outro lado, o algoritmo somente aplica as operações elementares com as linhas à matriz identidade I do lado direito de M. Assim, quando o algoritmo terminar, a matriz à direita de M, é igual a

$$E_k \cdots E_2 E_1 I = E_k \cdots E_2 E_1 = Q.$$

E por tomar $P = Q^t$, iremos obter $D = P^t A P$.

10.5 Classificação das Equações do 2º grau

Agora, vamos cumprir o prometido no início do capítulo e aplicar a teoria para classificar todas as formas quadráticas. Para isso, iniciemos com uma equação do 2º grau completa nas variáveis x, y da forma:

$$g(x,y) = b_{11}x^2 + b_{12}xy + b_{22}y^2 + a_1 x + a_2 y + a = 0.$$

Se fizermos

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}, A = \begin{bmatrix} b_{11} & \frac{b_{12}}{2} \\ \frac{b_{12}}{2} & b_{22} \end{bmatrix} \text{ e } A_1 = \begin{bmatrix} a_1 & a_2 \end{bmatrix},$$

então, poderemos escrever

$$\begin{aligned} g(x,y) &= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} b_{11} & \frac{b_{12}}{2} \\ \frac{b_{12}}{2} & b_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + a \\ &= \mathbf{x}^t A \mathbf{x} + A_1 \mathbf{x} + a = 0. \end{aligned}$$

Como A é uma matriz simétrica, pelo teorema 10.8, existe uma matriz invertível e ortogonal P tal que

$$\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = P^t A P \text{ e seja } \mathbf{y} = \begin{bmatrix} x' \\ y' \end{bmatrix} \text{ tal que } \mathbf{x} = P\mathbf{y}.$$

E a equação inicial tomará a forma

$$\begin{aligned} g(x',y') &= \mathbf{y}^t \left(P^t A P \right) \mathbf{y} + A_1 P \mathbf{y} + a \\ &= \begin{bmatrix} x' & y' \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} + \begin{bmatrix} d_1 & d_2 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} + a \\ &= \lambda_1 x'^2 + \lambda_2 y'^2 + d_1 x' + d_2 y' + a = 0, \end{aligned}$$

onde d_1 e d_2 serão os coeficientes obtidos ao fazer $A_1 P$. Voltando à equação, temos duas situações:

Se $\lambda_1\lambda_2 \neq 0$ (Se $\lambda_1\lambda_2 < 0$, temos uma hipérbole, e se $\lambda_1\lambda_2 > 0$, temos a elipse) podemos completar os quadrados, se necessário. E, com isso, obtemos

$$q(x', y') = \lambda_1 \left(x' + \frac{d_1}{2\lambda_1}\right)^2 + \lambda_2 \left(y' + \frac{d_2}{2\lambda_2}\right)^2 + b = 0,$$

onde $b = a - \left(\frac{d_1}{2\lambda_1}\right)^2 - \left(\frac{d_2}{2\lambda_2}\right)^2$. Assim, basta realizarmos uma translação no $\mathbb{R}^2$, dada por

$$x'' = x' + \frac{d_1}{2\lambda_1}, \qquad y'' = y' + \frac{d_2}{2\lambda_2},$$

para que a equação se torne

$$q(x'', y'') = \lambda_1 x''^2 + \lambda_2 y''^2 + b = 0.$$

Se $\lambda_1\lambda_2 = 0$ (parábola) Pelo menos um dos dois autovalores é não nulo, e digamos que $\lambda_1 \neq 0$ então, podemos completar o quadrado com respeito a x' e obtemos

$$q(x', y') = \lambda_1 \left(x' + \frac{d_1}{2\lambda_1}\right)^2 + d_2 y' + b' = 0,$$

onde $b' = a - \left(\frac{d_1}{2\lambda_1}\right)^2$. Fazendo a mudança de coordenadas

$$x'' = x' + \frac{d_1}{2\lambda_1}, \qquad y'' = y',$$

a equação se torna

$$q(x'', y'') = \lambda_1 x''^2 + d_2 y''^2 + b' = 0.$$

Em ambos os casos, é fácil classificar a equação. Além disso, podemos determinar quais mudanças de coordenadas foram necessárias para levá-la a essa configuração.

Percebemos, também, que esse procedimento se generaliza para qualquer número de variáveis ≥ 2.

Observação 10.20
Observe que no procedimento anterior, passamos da matriz

$$A = \begin{bmatrix} b_{11} & \frac{b_{12}}{2} \\ \frac{b_{12}}{2} & b_{22} \end{bmatrix} \text{ para a matriz } \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

por uma mudança de coordenadas. Utilizamos a matriz P, que é ortogonal, e obtemos

$$\det \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \det(A) = \frac{1}{4}\left(4b_{11}b_{22} - b_{12}^2\right).$$

Com exceção dos casos nos quais a equação representa figuras degeneradas, a classificação dessa equação em elipse, parábola ou hipérbole depende, apenas, do sinal das constantes λ_1 e λ_2, e podemos determinar esses sinais se conhecemos o produto $\lambda_1\lambda_2 = \det(A)$. Esse sinal pode ser obtido calculando $4b_{11}b_{22} - b_{12}^2$. Por isso, chamamos esse número de **discriminante da equação do 2º grau**.

Exemplo 10.21
Vamos aplicar o procedimento descrito acima para classificar a equação dada por

$$4x^2 - 24yx + 56x + 11y^2 - 58y + 95 = 0.$$

Inicialmente, podemos escrever

$$\begin{aligned} g(x,y) &= 4x^2 - 24yx + 11y^2 + 56x - 58y + 95 \\ &= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 4 & -12 \\ -12 & 11 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 56 & -58 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + 95 \\ &= \mathbf{x}^t A \mathbf{x} + A_1 \mathbf{x} + 95 = 0. \end{aligned}$$

Depois, precisamos calcular a base de autovetores de

$$A = \begin{bmatrix} 4 & -12 \\ -12 & 11 \end{bmatrix} \Rightarrow \Delta_A(x) = x^2 - 15x - 100 = (x+5)(x-20).$$

i) Calculando o autovetor associado a $\lambda_1 = -5$, obtemos $\left[\begin{smallmatrix} 4 \\ 3 \end{smallmatrix}\right]$.
ii) O autovetor associado a $\lambda_2 = 20$ é $\left[\begin{smallmatrix} -3 \\ 4 \end{smallmatrix}\right]$.

Portanto, em relação à base $\beta = \left\{ \left[\begin{smallmatrix} 4/5 \\ 3/5 \end{smallmatrix}\right], \left[\begin{smallmatrix} -3/5 \\ 4/5 \end{smallmatrix}\right] \right\}$, as coordenadas dos vetores serão escritas por ser $\mathbf{u} = \left[\begin{smallmatrix} x' \\ y' \end{smallmatrix}\right]$ e considerando a matriz

$$P = \begin{bmatrix} 4/5 & -3/5 \\ 3/5 & 4/5 \end{bmatrix}, \text{ então } \begin{bmatrix} x \\ y \end{bmatrix} = P \begin{bmatrix} x' \\ y' \end{bmatrix}.$$

E, no sistema de coordenadas β, a equação se torna

$$\begin{aligned} g(x',y') &= \begin{bmatrix} x' & y' \end{bmatrix} \begin{bmatrix} -5 & 0 \\ 0 & 20 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} + \begin{bmatrix} 10 & -80 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} + 95 \\ &= -5x'^2 + 20y'^2 + 10x' - 80y' + 95 \\ &= -5(x'^2 - 2x') + 20(y'^2 - 4y') + 95 \\ &= -5(x'^2 - 2x' + 1) + 20(y'^2 - 4y' + 4) + (95 + 5 - 80) \\ &= -5(x'-1)^2 + 20(y'-2)^2 + 20. \end{aligned}$$

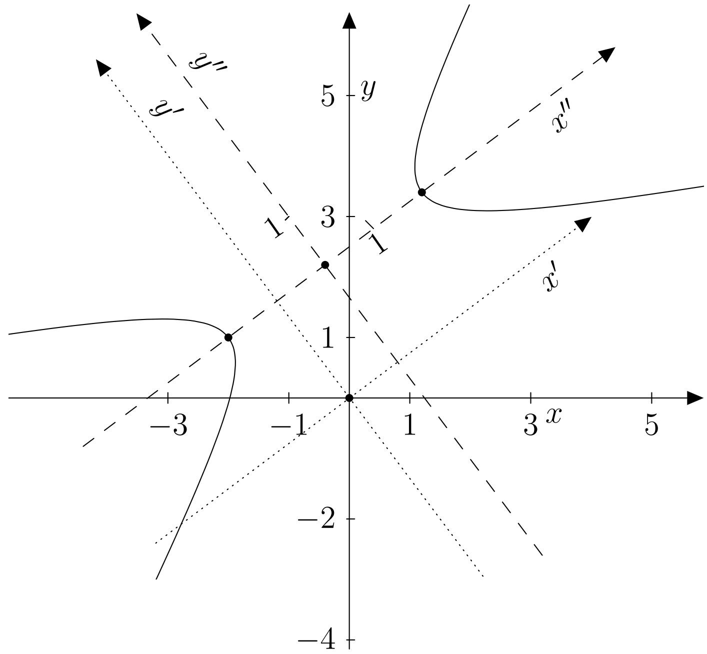
Fazendo a mudança de coordenadas

$$x'' = x' - 1, \quad y'' = y' - 2,$$

obtemos

$$g(x'',y'') = -5x''^2 + 20y''^2 + 20 = 0 \Leftrightarrow \frac{x''^2}{4} - y''^2 = 1.$$

Que é, portanto, uma hipérbole no sistema $x''y''$, que não cruza o eixo y''.



Vamos utilizar o segundo método (o algoritmo) para eliminar o fator xy na equação

$$4x^2 - 24yx + 56x + 11y^2 - 58y + 95 = 0$$
$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 4 & -12 \\ -12 & 11 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 56 & -58 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + 95 = 0.$$

Considere a matriz

$$M = \begin{bmatrix} 4 & -12 & 1 & 0 \\ -12 & 11 & 0 & 1 \end{bmatrix}.$$

Para escaloná-la, faça $\ell_2 \to \ell_2 + 3\ell_1$ e $\mathbf{c}_2 \to \mathbf{c}_2 + 3\mathbf{c}_1$ e obtemos

$$\begin{bmatrix} 4 & 0 & 1 & 0 \\ 0 & -25 & 3 & 1 \end{bmatrix}.$$

Tomando

$$P = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}, \text{ temos que } \begin{bmatrix} 4 & 0 \\ 0 & -25 \end{bmatrix} = P^t AP.$$

Portanto, se as coordenadas de $\mathbf{u}$ são $\left[\begin{smallmatrix} x' \\ y' \end{smallmatrix}\right]$, com respeito à base $\beta = \{\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 3 \\ 1 \end{smallmatrix}\right]\}$, então

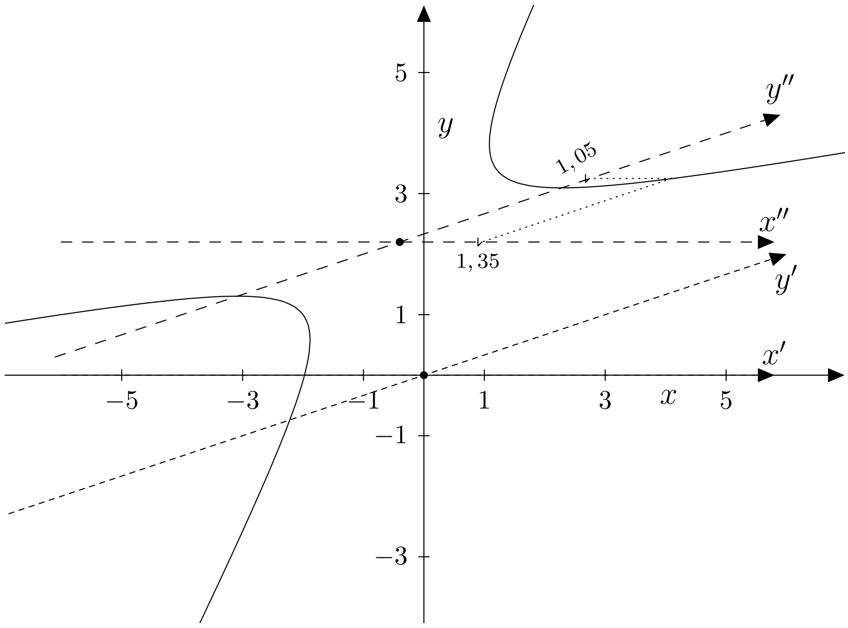
$$P = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}, \text{ e sabemos que } \begin{bmatrix} x \\ y \end{bmatrix} = P \begin{bmatrix} x' \\ y' \end{bmatrix}.$$

Então,

$$\begin{aligned} q(x', y') &= \begin{bmatrix} x' & y' \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & -25 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} + \begin{bmatrix} 56 & 110 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} + 95 \\ &= 4x'^2 - 25y'^2 + 56x' + 110y' + 95 \\ &= 4(x'^2 + 14x' + 49) - 25\left(y'^2 - \frac{22}{5}y' + \left(\frac{22}{5}\right)^2\right)^2 - 4 \times 49 + 25\left(\frac{22}{5}\right)^2 + 95 \\ &= 4(x' + 7)^2 - 25(y'^2 - \frac{11}{5})^2 + 20. \end{aligned}$$

Fazendo $x'' = x' + 7$ e $y'' = y' - \frac{11}{5}$, temos

$$4x''^2 - 25y''^2 + 20 = 0 \Leftrightarrow -\frac{x''^2}{5} + \frac{y''^2}{\frac{4}{5}} = 1.$$



Veja que o sistema de coordenadas $x''y''$ não é ortogonal nem preserva as medidas.

Com as técnicas propostas aqui, é possível classificar qualquer equação do segundo grau como uma das situações: degeneradas ou como

$$y^2 = 4px,\ \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,\ \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1.$$

Exercícios resolvidos

R10.1. Demonstre a proposição 10.3. Fixe um ponto F (Foco), uma linha ℓ (a diretriz) que não contém F, escolha, um número Real não-negativo e (a excentricidade). Considere os pontos P, satisfazendo

$$\operatorname{dist}(P, F) = e \operatorname{dist}(P, \ell).$$

Mostre que, se: $0 \leq e < 1$ obtemos uma elipse; se $e = 1$, obtemos uma parábola; e se $e > 1$, obtemos uma hipérbole.

Solução: Observe, que, se $e = 0$, claramente obtemos uma circunferência de raio 0 e centro P. Suponha, que $P = \left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$ e, $F = \left[\begin{smallmatrix} p \\ 0 \end{smallmatrix}\right]$ e ℓ seja o eixo do y, isto é, a reta $x = 0$. Pela definição, temos

$$\sqrt{(x-p)^2 + y^2} = e|x| \Leftrightarrow (1 - e^2)x^2 - 2px + y^2 + p^2 = 0.$$

Se $e = 1$, então, podemos reescrever a equação acima como

$$y^2 = 2p(x - \frac{p}{2}),$$

que é uma parábola.

Se $e \neq 1$, podemos dividir por $1 - e^2 \neq 0$. Completando o quadrado e dividindo por $-\frac{p^2}{1-e^2}$, obtemos

$$\frac{\left(x - \frac{p}{1-e^2}\right)^2}{\frac{p^2e^2}{(1-e^2)^2}} + \frac{y^2}{\frac{p^2e^2}{1-e^2}} = 1.$$

Se $e < 1$, então $1 - e^2 > 0$ e, com isso, temos uma elipse com

$$a^2 = \frac{p^2e^2}{(1-e^2)^2} \text{ e } b^2 = \frac{p^2e^2}{1-e^2}.$$

Se $e > 1$, então $1 - e^2 < 0$ e, com isso, temos uma hipérbole. □

R10.2. Prove o teorema 10.4. Sejam $\langle\,,\,\rangle$ um produto interno em $\mathbb{R}^n$ e $T : \mathbb{R}^n \to \mathbb{R}^n$ um operador linear. Então, T será um operador autoadjunto se, e somente se,

$$\langle T\mathbf{u}, \mathbf{v}\rangle = \langle \mathbf{u}, T\mathbf{v}\rangle\,, \text{ para todo } \mathbf{u}, \mathbf{v} \in \mathbb{R}^n.$$

Solução: Seja $\alpha = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ a base ortonormal de $\mathbb{R}^n$, na qual a matriz $[T]^\alpha_\alpha = [a_{ij}]$ é simétrica. Os coeficientes a_{ij} são determinados por

$$T(\mathbf{u}_j) = \sum_{i=1}^{n} a_{ij}\mathbf{u}_i, \text{ para todo } 1 \leq j \leq n.$$

Por calcular

$$\begin{aligned}
\langle T(\mathbf{u}_j), \mathbf{u}_k\rangle &= \left\langle \sum_{i=1}^{n} a_{ij}\mathbf{u}_i, \mathbf{u}_k \right\rangle \\
&= a_{kj} = a_{jk} \\
&= \left\langle \mathbf{u}_j, \sum_{i=1}^{n} a_{ik}\mathbf{u}_i \right\rangle = \langle \mathbf{u}_j, T(\mathbf{u}_k)\rangle\,.
\end{aligned}$$

E, com isso, vemos que essa propriedade é verdadeira se substituírmos os vetores $\mathbf{u}, \mathbf{v}$ pelos vetores da base α. Seja agora $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ dois vetores quaisquer, e por escrevê-los em termos da base α, obtemos $\mathbf{u} = \sum_{i=1}^{n} x_i \mathbf{u}_i$ e $\mathbf{v} = \sum_{j=1}^{k} y_j \mathbf{u}_j$ e, portanto,

$$\begin{aligned}
\langle T\mathbf{u}, T\mathbf{v}\rangle &= \left\langle T(\sum_{i=1}^{n} x_i \mathbf{u}_i), T(\sum_{j=1}^{k} y_j \mathbf{u}_j)\right\rangle \\
&= \sum_{i=1}^{n} x_i \left\langle T(\mathbf{u}_i), T(\sum_{j=1}^{k} y_j \mathbf{u}_j)\right\rangle \\
&= \sum_{i=1}^{n}\sum_{j=1}^{k} x_i y_j \left\langle T(\mathbf{u}_i), T(\mathbf{u}_j)\right\rangle \\
&= \sum_{i=1}^{n}\sum_{j=1}^{k} x_i y_j \left\langle \mathbf{u}_i, \mathbf{u}_j\right\rangle \\
&= \sum_{i=1}^{n} x_i \left\langle \mathbf{u}_i, \sum_{j=1}^{k} y_j \mathbf{u}_j\right\rangle \\
&= \left\langle \sum_{i=1}^{n} x_i \mathbf{u}_i, \sum_{j=1}^{k} y_j \mathbf{u}_j\right\rangle = \langle \mathbf{u}, \mathbf{v}\rangle .
\end{aligned}$$

Reciprocamente, suponha que:

$$\langle T\mathbf{u}, \mathbf{v}\rangle = \langle \mathbf{u}, T\mathbf{v}\rangle \text{ para todo } \mathbf{u}, \mathbf{v} \in \mathbb{R}^n.$$

E seja $\beta = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ uma base ortonormal qualquer de $\mathbb{R}^n$, então, a matriz $[T]_\beta^\beta = [b_{ij}]$ será determinada por satisfazer

$$T(\mathbf{v}_j) = \sum_{i=1}^{n} b_{ij}\mathbf{v}_i.$$

E como β é uma base ortonormal temos que

$$b_{lj} = \langle T(\mathbf{v}_j), \mathbf{v}_l\rangle = \langle \mathbf{v}_j, T(\mathbf{v}_l)\rangle = b_{jl}.$$

Portanto, $[T]_\beta^\beta$ é simétrica. Daí, segue que, se a matriz de uma transformação é simétrica, então ela é simétrica, também, com relação a qualquer base ortonormal. $\square$

R10.3. Prove a proposição 10.6. Seja $\langle\, , \rangle$ um produto interno em $\mathbb{R}^n$. Se $T : \mathbb{R}^n \to \mathbb{R}^n$ é um operador autoadjunto. Então:

a) Autovetores correspondentes a autovalores distintos são ortogonais.

b) T sempre possui pelo menos um autovalor.

Solução: a) Sejam $i \neq j$ quaisquer:

$$\begin{aligned}(\lambda_i - \lambda_j)\langle \mathbf{u}_i, \mathbf{u}_j\rangle &= \langle \lambda_i \mathbf{u}_i, \mathbf{u}_j\rangle - \langle \mathbf{u}_i, \lambda_j \mathbf{u}_j\rangle \\ &= \langle T\mathbf{u}_i, \mathbf{u}_j\rangle - \langle \mathbf{u}_i, T\mathbf{u}_j\rangle = \langle T\mathbf{u}_i, \mathbf{u}_j\rangle - \langle T\mathbf{u}_i, \mathbf{u}_j\rangle = 0,\end{aligned}$$

uma vez que T é autoadjunto. Como $\lambda_i - \lambda_j \neq 0$ e $(\lambda_i - \lambda_j)\langle \mathbf{u}_i, \mathbf{u}_j\rangle = 0$, isso implica que $\langle \mathbf{u}_i, \mathbf{u}_j\rangle = 0$.

b) Considere a aplicação $\phi : \mathbb{R}^n \to \mathbb{R}$, definida por $\phi(\mathbf{u}) = \langle T\mathbf{u}, \mathbf{u}\rangle$, que é contínua e deverá assumir um máximo em $K : \{\mathbf{u} \in \mathbb{R}^n : \|\mathbf{u}\| = 1\}$, uma vez que o mesmo é compacto. Seja $\mathbf{u}_1$ esse ponto, e λ_1 o valor de ϕ nesse ponto. Então, para todo $\mathbf{u} \in \mathbb{R}^n$, vale

$$\begin{aligned}\langle (T - \lambda_1 I)\mathbf{u}, \mathbf{u}\rangle &= \|\mathbf{u}\|^2 \left\langle (T-\lambda_1 I)\frac{1}{\|\mathbf{u}\|}\mathbf{u}, \frac{1}{\|\mathbf{u}\|}\mathbf{u}\right\rangle \\ &= \|\mathbf{u}\|^2 \left[\left\langle T\left(\frac{1}{\|\mathbf{u}\|}\mathbf{u}\right), \frac{1}{\|\mathbf{u}\|}\mathbf{u}\right\rangle - \lambda_1\right] \leq 0,\end{aligned}$$

mostrando que

$$\langle (T - \lambda_1 I)\mathbf{u}, \mathbf{u}\rangle \leq 0, \ \forall \mathbf{u} \in \mathbb{R}^n.$$

Então, se escolhermos $\mathbf{u} = \mathbf{u}_1 + t\mathbf{v}$, com $t \in \mathbb{R}$ e $\mathbf{v} \in \mathbb{R}^n$, teremos

$$\begin{aligned}&\langle (T - \lambda_1 I)(\mathbf{u}_1 + t\mathbf{v}), \mathbf{u}_1 + t\mathbf{v}\rangle \\ &\qquad = t^2 \langle (T-\lambda_1 I)(\mathbf{v}), \mathbf{v}\rangle + t\left[\langle (T-\lambda_1 I)(\mathbf{u}_1, \mathbf{v}\rangle + \langle (T-\lambda_1 I)\mathbf{v}, \mathbf{u}_1\rangle\right] \\ &\qquad\qquad + \langle (T-\lambda_1 I)\mathbf{u}_1, \mathbf{u}_1\rangle \\ &\qquad = t^2 \langle (T-\lambda_1 I)(\mathbf{v}), \mathbf{v}\rangle + 2t\langle (T-\lambda_1 I)\mathbf{u}_1, \mathbf{v}\rangle \leq 0.\end{aligned}$$

Temos uma equação do 2º grau com o coeficiente que acompanha t^2 negativo e, portanto, o discriminante precisa ser menor ou igual a zero, isto é, $\langle (T-\lambda_1 I)\mathbf{u}_1, \mathbf{v}\rangle = 0$ para todo $\mathbf{v} \in \mathbb{R}^n$. Então $(T - \lambda_1 I)\mathbf{u}_1 = 0$, mostrando, assim, a existência de um autovalor λ_1 e de um autovetor $\mathbf{u}_1$. $\square$

R10.4. Mostre que a equação da reta tangente à parábola $y^2 = 4px$ em $P = \left[\begin{smallmatrix} x_0 \\ y_0 \end{smallmatrix}\right]$ é

$$y_0 y = 2p(x + x_0).$$

Solução: Queremos determinar a equação da reta tangente

$$y - y_0 = m(x - x_0)$$

à parábola passando por $P = \left[\begin{smallmatrix} x_0 \\ y_0 \end{smallmatrix}\right]$. Precisamos determinar m. Para isto, faça a substituição $y = mx - mx_0 + y_0$ na equação da parábola e teremos

$$(mx - mx_0 + y_0)^2 = 4px \Leftrightarrow m^2x^2 + \left(2my_0 - 2m^2x_0 - 4p\right)x + m^2x_0^2 - 2mx_0y_0 + y_0^2.$$

Essa equação do 2º grau só pode ter uma solução e, portanto, o seu discriminante precisa ser igual a zero, isto é,

$$\left(2my_0 - 2m^2x_0 - 4p\right)^2 - 4m^2\left(m^2x_0^2 - 2mx_0y_0 + y_0^2\right) = 0.$$

Fazendo as contas obtemos

$$16p(x_0m^2 - y_0m + p) = 0 \Leftrightarrow x_0m^2 - y_0m + p = 0.$$

Resolvendo para m, obtemos

$$m = \frac{y_0 + \sqrt{y_0^2 - 4x_0p}}{2x_0}.$$

Isso nos dá $m = \frac{y_0}{2x_0}$ visto que P pertence à parábola. Substituindo m na equação da reta inicial, obtemos a equação enunciada. $\square$

R10.5. Prove o teorema Espectral 10.8: Seja $\langle\,,\,\rangle$ um produto interno em $\mathbb{R}^n$. Se $T : \mathbb{R}^n \to \mathbb{R}^n$ é um operador auto-adjunto, então existe uma base ortogonal de $\mathbb{R}^n$ composta por autovetores de T.

Solução: Vamos demonstrar por indução sobre $n = \dim \mathbb{R}^n$. No caso que $n = 1$, todo operador é autoadjunto, uma vez que toda matriz 1×1 é simétrica. Suponha o resultado verdadeiro em espaço com dimensão $n - 1$, e vamos provar quando a dimensão é n. Seja $T : R^n \to \mathbb{R}^n$ um operador autoadjunto, pela proposição 10.6 b., existe um autovetor unitário $\mathbf{u}_1$ associado ao autovalor λ_1. Considere o subespaço $W = \operatorname{Span}\{\mathbf{u}_1\}$, que é um subespaço T-invariante de dimensão 1. Segue, pelo lema 10.7, que o subespaço, com dimensão $n-1$, $W^\perp$ também é T-invariante. Logo, $\bar{T} = T|_{W^\perp} : W^\perp \to W^\perp$ também é autoadjunto, pela hipótese de indução, existe uma base ortonormal $\{\mathbf{u}_2, \dots, \mathbf{u}_n\} \subset W^\perp$ formada por autovetores de $\bar{T}$. Segue que $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ é uma base ortonormal formada por autovetores de T. $\square$

R10.6. Seja $Q : \mathbb{R}^n \to \mathbb{R}$ uma forma quadrática associada à forma bilinear simétrica f. Mostre que vale a seguinte identidade:

$$f(\mathbf{u}, \mathbf{v}) = \frac{1}{2}\left[Q(\mathbf{u} + \mathbf{v}) - Q(\mathbf{u}) - Q(\mathbf{v})\right],$$

conhecida como fórmula de polarização.

Solução: Sabemos que $Q(\mathbf{u}) = f(\mathbf{u}, \mathbf{u})$, e seja $\mathbf{u}, \mathbf{w}$ e $\mathbf{v} \in \mathbb{R}^n$ e $\alpha \in \mathbb{R}$. Então

$$\begin{aligned}
Q(\mathbf{u}+\mathbf{v}) - Q(\mathbf{u}) - Q(\mathbf{v}) &\\
&= f(\mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v}) - f(\mathbf{u}, \mathbf{v}) - f(\mathbf{v}, \mathbf{v})\\
&= f(\mathbf{u}, \mathbf{u}+\mathbf{v}) + f(\mathbf{v}, \mathbf{u}+\mathbf{v}) - f(\mathbf{u}, \mathbf{v}) - f(\mathbf{v}, \mathbf{v})\\
&= f(\mathbf{u}, \mathbf{u}) + f(\mathbf{u}, \mathbf{v}) + f(\mathbf{v}, \mathbf{u}) + f(\mathbf{v}, \mathbf{v}) - f(\mathbf{u}, \mathbf{v}) - f(\mathbf{v}, \mathbf{v})\\
&= 2f(\mathbf{u}, \mathbf{v}).
\end{aligned}$$

Dividindo por 2, obtemos a fórmula de polarização. $\square$

Exercícios propostos

P10.1. Identifique a figura e ache sua posição quando a sua equação for:

a) $2xy + 3x - y + 1 = 0$;

b) $x^2 + y^2 + xy - x + 1 = 0$;

c) $9x^2 + 12y^2 + 9z^2 - 6xy - 6yz - 1 = 0$;

d) $11x^2 + 10y^2 + 6z^2 - 12xy - 8yz + 4xz - 12 = 0$;

e) $xy + x + y = 0$;

f) $3x^2 - 4\sqrt{3}xy - y^2 + 20y - 25 = 0$;

g) $4xy + 3y^2 + 2\sqrt{5}x + 4\sqrt{5}y = 0$;

h) $9x^2 + 16y^2 + 25z^2 + 24xy - 40x + 30y = 0$.

P10.2. Mostre que as equações paramétricas $x = a\cos t$ e $y = b\,\mathrm{sen}\,t$, $a > 0$, $b > 0$ e $0 \leq t < 2\pi$ definem uma elipse cuja equação em coordenadas cartesianas é $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$.

P10.3. Mostre que as equações paramétricas $x = t$ e $y = \frac{t^2}{4p}$, com $t \in \mathbb{R}$, definem uma parábola cuja equação em coordenadas cartesianas é $y^2 = 4px$.

P10.4. Seja $\cosh t = \frac{e^x + e^{-x}}{2}$ e $\mathrm{senh} = \frac{e^x - e^{-x}}{2}$. Mostre que as equações paramétricas $x = a\cosh t$ e $y = b\,\mathrm{senh}\,t$, $a > 0$, $b > 0$ e $t \in \mathbb{R}$ definem uma hipérbole cuja equação em coordenadas cartesianas é $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$.

P10.5. Para as matrizes abaixo, aplique o algoritmo para encontrar uma matriz invertível P, tal que $D = P^tAP$ seja diagonal, onde

$$A = \begin{bmatrix} 1 & -3 & 2 \\ -3 & 7 & -5 \\ 2 & -5 & 8 \end{bmatrix} \text{ e } A = \begin{bmatrix} 4 & -5 & 7 \\ -5 & -6 & 8 \\ 7 & 8 & -9 \end{bmatrix}.$$

P10.6. a) Dê a transformação linear que descreve o movimento rígido que leva o segmento de extremos $\left[\begin{smallmatrix} -6 \\ 2 \end{smallmatrix}\right]$ e $\left[\begin{smallmatrix} -1 \\ 2 \end{smallmatrix}\right]$ ao segmento de extremos $\left[\begin{smallmatrix} -2 \\ 6 \end{smallmatrix}\right]$ e $\left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$, respectivamente.

b) Mostre que essa transformação é uma rotação, e encontre o seu ângulo.

P10.7. Mostre que a equação da reta tangente à elipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ em $P = \left[\begin{smallmatrix} x_0 \\ y_0 \end{smallmatrix}\right]$ é

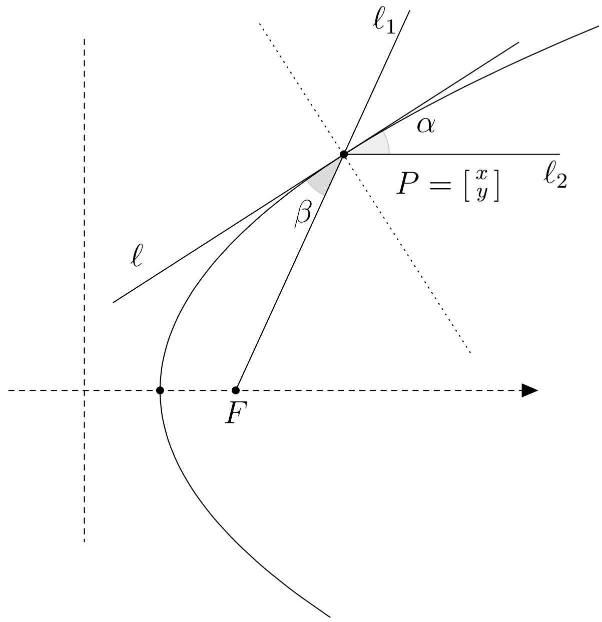
$$b^2x_0x + a^2y_0y = a^2b^2.$$

P10.8. Suponha que sejam dadas as retas $y = m_1x + k_1$ e $y = m_2x + k_2$, as quais se interceptam em um ponto P. Seja θ o ângulo entre elas. Então,

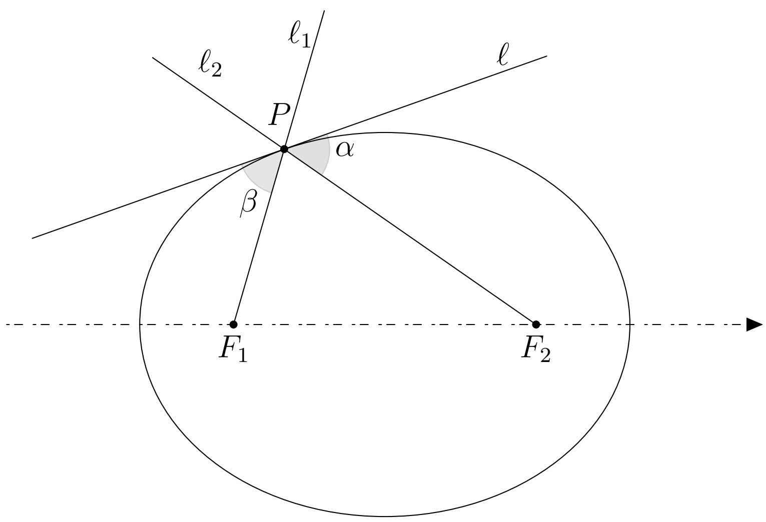
$$\mathrm{tg}\,\theta = \frac{m_2 - m_1}{1 + m_1m_2}, \;\; m_1m_2 \neq -1,$$

onde m_1 é o coeficiente da reta onde se inicia o ângulo.

P10.9. Seja P um ponto de uma parábola e ℓ a reta tangente à parábola em P. Sejam ℓ_1 a reta passando pelo foco F e o ponto P e ℓ_2 a reta que passa por P, e que é paralela ao eixo x, então os ângulos, no ponto P, α entre ℓ e ℓ_1 e β entre ℓ e ℓ_2, são iguais.



P10.10. Sejam P um ponto sobre uma elipse e ℓ a reta tangente a esta em P. Considere ℓ_1 a reta ligando o foco em F_1 e ℓ_2, a reta ligando o foco F_2 a P. Então, os ângulos: α, entre as retas ℓ e ℓ_1, e β, entre ℓ e ℓ_2, são iguais.



Bibliografia

[1] Paul R. Halmos, *Finite-Dimensional Vector Spaces*. Springer-Verlag, 1974.

[2] BIBLIOGRAFIA AINDA ESTÁ POR FAZER…