

Álgebra linear algorítmica

S. C. Coutinho

Prefácio

Agradeço a todos os alunos que cursaram *álgebra linear algorítmica* entre 2010 e 2021 e que serviram de cobaias para a disciplina e para as notas que a acompanham, especialmente Fabio Ferman, Fillipe Barros da Silva, João Augusto Marrara Marzagão, Raul Barbosa, Mateus Gregório, Rochanne de Miranda Corrêa, Filipe Qiang Zhou, Júlio Zynger, Edberg dos Santos Franco, Victor Lima Campos, Daniel de Souza Gonçalves, José Alexandre Siqueira Junior, João Felipe C. da Silva, Carlos Eduardo da Silva Martins, Yuri de Jesus Lopes de Abreu, Jean Carlos Silva De Souza, João Vitor de Oliveira Silva, Igor Carpanese, Lucca Martins Felix, Bruno Neves Saraiva, Thalles Nonato, Sabrina Alencar das Neves, Vinícius Lettieri, Daniel Kiyoshi Hashimoto Vouzella de Andrade, Matheus Moura Gorchiinsky, Luan Martins, Felipe Dias de Melo, Carlos Eduardo Valladares da Mota, Sabrina Neves, Lorena Mamede Botelho Bruno Llacer Trott, Gustavo Martins Ferreira de Aquino, Tales Coutinho Moreira, Raphael Felipe de Cantuária Mesquita, Matheus Augusto Vargas da Silva, Jorge Luis Ribeiro de Oliveira, Jefferson Maxwell de Oliveira Ferreira, João Henrique Schmidt De Carvalho, Gabriele Jandres Cavalcanti, Carlos Henrique Bravo Serrado, Daniel Corcino de Albuquerque, Robson Duarte, Daniel Li Vam Mam, Pedro Poppolino, Helena Serrano Cardoso da Costa e Laura Serrano Cardoso da Costa, Michel Monteiro Schorr, Bruno de Queiroz Cauper Oliveira, Gabriel Oliveira de Março, Ana Cláudia Ribeiro dos Santos, Renato Sousa do Nascimento, Gabriel Ferreira Leão, Raisa Christina Nascimento Gonçalves, Lucas de Castro Carelli, Carlos Eduardo de Schuller Banjar, Gabriel Duarte Soares, Matheus do Santos Tibúrcio, Gustavo Henrique Lemos de Lima, que detectaram e ajudaram a corrigir alguns dos inúmeros erros do manuscrito original. Certamente muitos outros erros ainda permanecem, por favor me ajude a corrigi-los enviando uma mensagem para collier@ic.ufrj.br. Como todo mundo que já leu um contrato sabe, a parte mais importante está escrita em letra miúda. Neste livro isto aparece claramente marcado:



Não deixe de ler a letra miúda marcada desta maneira maneira porque se trata de algo importante!

Finalmente, a epígrafe é o refrão da música “For the love” de Jamie Cullum.

Every time you feel like giving up,
Every time you'd swear you'd had enough,
Just remember what you're dreaming of,
But when you do it, you do it for the love,
Yeah when you do it, you do it for the love.

Jamie Cullum

Para Adilson e Astréa

Sumário

Prefácio	iii
	v
Parte 1. Álgebra linear no plano	1
Capítulo 1. Vetores e operadores lineares do plano	3
1. Vetores	3
2. Operadores lineares	13
3. Matriz de um operador linear	21
Exercícios	32
Capítulo 2. Autovalores e autovetores	37
1. Autovalores e autovetores	37
2. Calculando autovalores e autovetores	42
3. O teorema espectral	47
Exercícios	64
Parte 2. Ferramentas	67
Capítulo 3. Matrizes	69
1. Definição e exemplos	69
2. Operações com matrizes	73
3. Funções de matrizes	81
Capítulo 4. Sistemas lineares	85
1. Sistemas com duas variáveis	85
2. Sistemas triangulares	88
3. Eliminação gaussiana	94
4. Aplicações	106
Exercícios	108
Capítulo 5. Determinante e matrizes inversas	111
1. Determinantes	111
2. Decomposição de matrizes	118
Exercícios	128

Parte 3. Espaços vetoriais e transformações lineares	131
Capítulo 6. Espaços multidimensionais	133
1. Sistemas homogêneos	133
2. O espaço \mathbb{R}^n	140
3. Subespaços	145
4. Bases e dimensão	149
5. Bases ortonormais	158
6. Operações com subespaços	168
Exercícios	179
Capítulo 7. Transformações lineares	185
1. Definição e exemplos	185
2. Núcleo e imagem	194
3. Mudança de base	200
Exercícios	214
Capítulo 8. Diagonalização de operadores	221
1. Autovalores e autovetores	221
2. Teorema espectral	229
3. Rotações	236
4. Busca na rede	249
Exercícios	255
Referências Bibliográficas	261

Parte 1

Álgebra linear no plano

Na primeira parte do curso estudamos todos os principais conceitos e resultados a serem introduzidos no curso no contexto do plano. Isto nos permite tratar o assunto de maneira mais concreta, usando ilustrações e escrevendo explicitamente fórmulas que, na terceira parte, serão tratadas de maneira mais abstrata.

CAPÍTULO 1

Vetores e operadores lineares do plano

Nossa meta neste capítulo consiste em introduzir vetores e operadores lineares do plano e sua representação em termos de matrizes.

1. Vetores

Um vetor é, essencialmente, um segmento de reta orientado e, como tal, tem:

- um comprimento, geralmente chamado de *módulo* ou *norma*;
- uma *direção*, dada pela reta subjacente ao segmento;
- um *sentido*, que nos diz para que lado da reta subjacente o segmento aponta.

Além disso suporemos que vetores não podem “flutuar” por onde desejarem. Fixaremos para todo o sempre um ponto do plano, que chamaremos de *origem* e denotaremos por O . Todos os vetores terão uma de suas extremidades na origem e a orientação do segmento será sempre da origem para a outra extremidade, como mostra a figura.



Designaremos vetores por letras, sem a necessidade de adicionar a tradicional seta no alto da letra. Se u for um vetor, seu módulo será denotado por $\|u\|$. Reservaremos as barras simples para o módulo de um número real; isto é, se $r \in \mathbb{R}$, então

$$|r| = \begin{cases} r & \text{se } r \geq 0; \\ -r & \text{se } r < 0; \end{cases}$$

Muitas das fórmulas que aparecem neste livro serão descritas apenas no caso em que o vetor é *unitário*; isto é, sua norma é igual a um. Porém, dado um vetor não nulo v do plano, podemos encontrar um vetor unitário que tem a mesma direção e sentido de v simplesmente

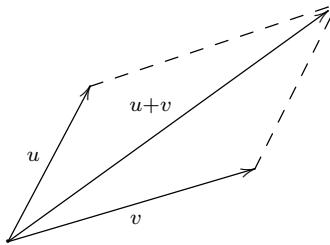
dividindo v por sua norma. Dizemos que o vetor

$$u = \frac{v}{\|v\|}$$

construído desta maneira é a *normalização* do vetor v .

1.1. Operações com vetores. Ao contrário dos segmentos de retas, vetores não são estáticos, podemos operar com eles. A operação mais simples é a soma de vetores, definida pela *regra do paralelogramo*:

dados dois vetores u e v , formamos o paralelogramo, com vértice na origem e lados u e v ; a soma $u + v$ corresponde à diagonal maior do paralelogramo orientada da origem para o vértice oposto; como mostra a figura.



A ideia que levou a esta definição é antiga e muito natural. Por exemplo, dois cavalos puxando uma balsa, um em cada lado de um canal, oferecem uma ilustração perfeita desta regra acima. Em seu famoso *Principia*, Newton prova a regra do paralelogramo no corolário I da Lei II, ilustrado na figura 1. Contudo, apesar de usar grandezas que hoje expressamos usando vetores, Newton não introduziu o conceito de vetor, que só foi formalizado no século XIX.

Corol. I.

Corpus viribus conjunctis diagonalem parallelogrammi eodem tempore describere, quo latera separatis.

Si corpus dato tempore, vi sola M , ferretur ab A ad B , & vi sola N , ab A ad C , compleatur parallelogramnum $ABDC$, & vi utraq; feretur id eodem tempore ab A ad D . Nam quoniam vis N agit secundum lineam AC ipsi BD parallelam, hæc vis nihil mutabit velocitatem acce-

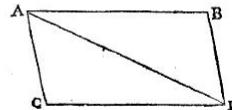
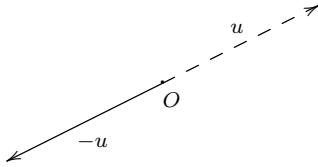


FIGURA 1. O corolário I do Principia

Já para subtrair o vetor u do vetor v , somamos a v o vetor $-u$, obtido invertendo-se o sentido da seta de u . Como todos os vetores têm que ter seu ponto de partida na

origem, uma maneira mais precisa de descrever esta receita consiste em dizer que, sobre a mesma reta ao longo da qual está u , desenhamos $-u$ como o segmento orientado de mesmo comprimento que u , mas que aponta no sentido oposto a u , como ilustra a figura.



Observe que, literalmente falando, não podemos aplicar a regra do paralelogramo a estes dois vetores. Afinal, eles são colineares e, por isso, não constituem os lados de um paralelogramo. Interpretaremos isto como significando que a soma destes vetores é o vetor zero, aquele que tem início e fim na origem, e que denotaremos por 0 . Sob estas convenções é fácil, mas muito monótono, verificar geometricamente as seguintes propriedades da soma e da subtração de vetores. Se u, v e w são vetores do plano, então

- $(u + v) + w = u + (v + w)$;
- $u + v = v + u$;
- $u + 0 = u$;
- $u + (-u) = 0$.

Segundo a primeira das propriedades acima, o posicionamento dos parêntesis não afeta o resultado final da adição de vetores. Com isso, se $k \in \mathbb{N}$, podemos abreviar

$$\underbrace{u + \cdots + u}_{k \text{ vezes}}$$

por ku , como de praxe. Como definimos $-u$ como sendo o vetor colinear e de sentido oposto a u , convém dizer que $(-1) \cdot u = -u$. Portanto, se k é um inteiro negativo, teremos

$$(1) \quad k \cdot u = \underbrace{-u - \cdots - u}_{|k| \text{ vezes}}.$$

Na verdade, vamos generalizar estas definições de modo a permitir o produto de qualquer número real λ por um vetor u . Para isso, declaramos λu como sendo o segmento orientado colinear a u cujo comprimento é igual a $|\lambda|$ vezes o comprimento de u . Para que esta definição seja compatível com (1), precisamos que λu tenha o mesmo sentido que u se $\lambda > 0$ e o sentido oposto se $\lambda < 0$. E quando o escalar é o zero? Pela regra anterior, o vetor obtido multiplicando o escalar 0 por um vetor u tem norma $0 \cdot \|u\| = 0$; de modo

que tem que ser o vetor nulo. O produto de um escalar por um vetor satisfaz as seguintes propriedades:

- $1 \cdot u = u$;
- $0 \cdot u = 0$;
- $\lambda(u + v) = \lambda u + \lambda v$;
- $(\lambda + \mu)u = \lambda u + \mu u$;
- $(\lambda\mu)u = \lambda(\mu u)$;

em que u, v e w são vetores do plano e $\lambda, \mu \in \mathbb{R}$. Note que em $0 \cdot u = 0$ o zero que multiplica u é um escalar, ao passo que o zero à direita do sinal de igualdade é o vetor nulo.

Como aplicação do que fizemos até aqui descreveremos a equação vetorial de uma reta r . Só precisamos considerar o caso em que r passa pela origem, porque em álgebra linear estudamos apenas subconjuntos (retas, planos, hiperplanos,...) que contém a origem. No caso da reta, basta escolher um vetor não nulo u ao longo de r , porque os demais vetores de r serão os múltiplos do vetor u ; isto é,

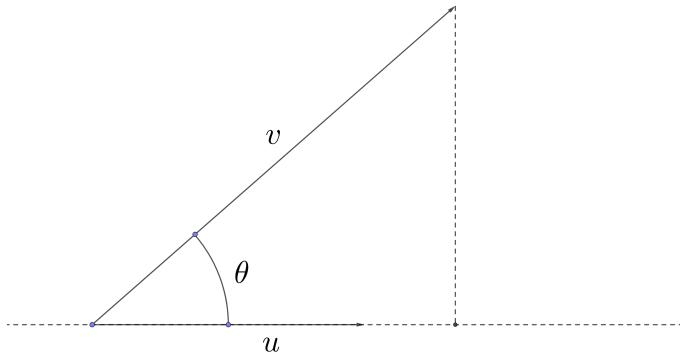
$$(2) \quad r = \{\lambda \cdot u \mid \lambda \in \mathbb{R}\}.$$

Neste caso dizemos que estamos considerando u como *vetor diretor* da reta r . Note que *qualquer vetor não nulo de r pode ser usado como vetor diretor desta reta*. Uma pergunta razoável é: de que forma a equação vetorial se relaciona à equação cartesiana da reta, que é aquela que aprendemos no ensino médio? Para respondê-la, precisamos introduzir coordenadas nos nossos vetores.

 Enquanto no ensino fundamental e médio uma reta é descrita como um conjunto de pontos, em álgebra linear a mesma reta corresponde a um conjunto de vetores. Naturalmente as duas linguagens se equivalem, pois ao ponto P da reta r corresponde o vetor \overrightarrow{OP} que vai da origem a P .

1.2. Projeções e bases ortonormais. Como ilustrado na figura abaixo, um exercício simples de trigonometria mostra que, projetando o segmento correspondente a um vetor v sobre a reta suporte do vetor u , obtemos um segmento de comprimento $\|v\| \cdot |\cos \theta|$, em que θ é o *menor* ângulo entre os vetores u e v . Usando isto, definimos a projeção do vetor v sobre o vetor u , como sendo o vetor $\text{Proj}_u(v)$, que tem comprimento $\|v\| |\cos \theta|$ e mesma reta suporte que u . O sentido da projeção é o mesmo de u se o ângulo θ for agudo, e oposto a u se θ for obtuso.

Naturalmente, podemos determinar se o ângulo θ é agudo ou obtuso a partir do cosseno. No primeiro caso, $\cos \theta$ é positivo; no segundo, negativo. Mas isto significa que se u for



um vetor unitário, então o vetor

$$(\|v\| \cos \theta) \cdot u$$

é colinear a u e tem o mesmo comprimento e sentido de $\text{Proj}_u(v)$; de modo que estes dois vetores são iguais. Naturalmente, quando u não for unitário, devemos substituí-lo por sua normalização. Portanto, em geral,

$$(3) \quad \text{Proj}_u(v) = \|v\| \cos \theta \frac{u}{\|u\|}.$$

A noção de projeção nos permite introduzir coordenadas para vetores do plano. Já vimos que, para descrever vetores, precisamos fixar o ponto que lhes serve de origem. Para introduzir coordenadas, fixamos também dois vetores unitários perpendiculares no plano, que denotaremos por e_1 e e_2 . O conjunto $\{e_1, e_2\}$ é conhecido como uma *base ortonormal* do plano. O adjetivo ortonormal é usado para lembrar que os dois vetores da base devem ser

- ortogonais, do grego para ângulo reto, significa o mesmo que perpendicular;
- normalizados, de modo que ambos têm norma igual a um.

Suponhamos que escolhemos uma base ortonormal $\varepsilon = \{e_1, e_2\}$, que estará fixa pelo resto deste artigo. Se v é um vetor do plano e ϕ é o *ângulo orientado* entre v e e_1 , então um argumento trigonométrico simples mostra que

$$(4) \quad v = (\|v\| \cos \phi) \cdot e_1 + (\|v\| \sin \phi) \cdot e_2;$$

como ilustrado na figura 2. Os números $\|v\| \cos \phi$ e $\|v\| \sin \phi$ são as *coordenadas* de v relativamente à base $\varepsilon = \{e_1, e_2\}$.

 A mudança de θ para ϕ na representação do ângulo foi feita porque, em geral, θ e ϕ *denotam ângulos diferentes*. Enquanto θ denota o *menor* ângulo entre v e e_1 (ilustrado na figura 3), ϕ

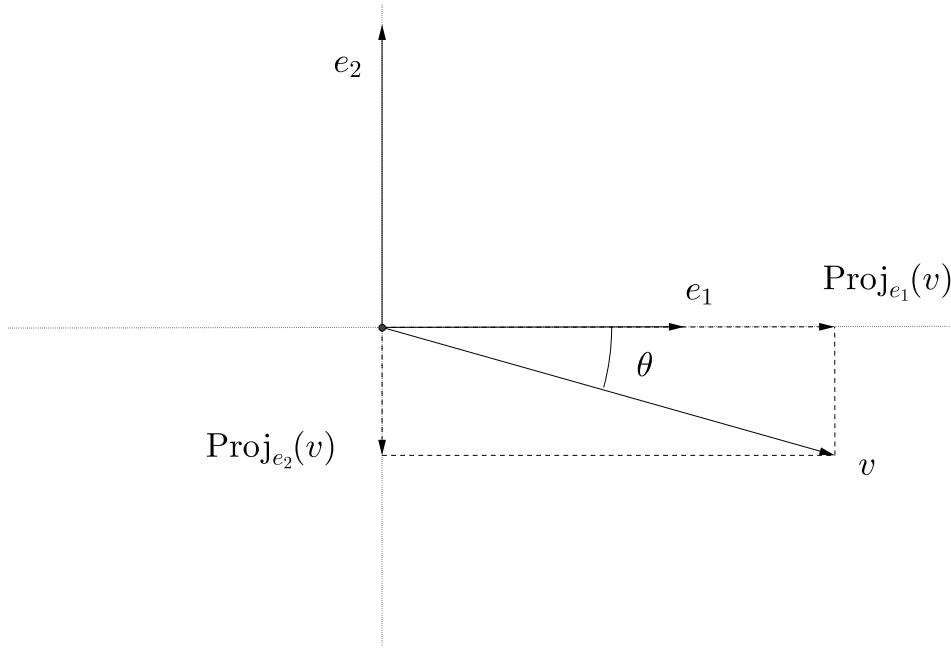
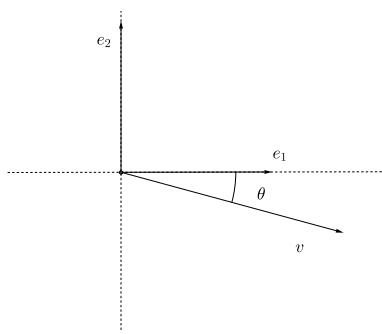
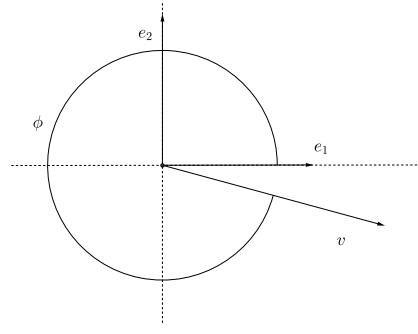


FIGURA 2. Projeções de um vetor em uma base ortonormal

FIGURA 3. Menor ângulo entre v e e_1 .FIGURA 4. Ângulo orientado entre v e e_1 .

corresponde ao ângulo *orientado* entre estes vetores (ilustrado na figura 4), que é o ângulo medido em trigonometria.

Só explicitaremos as coordenadas de um vetor em termos de sua norma e do ângulo orientado, quando isto for relevante para nossos propósitos. Em geral, escreveremos

$$v = (a_1, a_2), \quad \text{como abreviação de} \quad v = a_1 e_1 + a_2 e_2.$$

Isto só é possível porque, *uma vez que uma base ortonormal tenha sido fixada*, a cada vetor corresponde um único par de coordenadas. Para provar isto, digamos que

$$(5) \quad v = a_1e_1 + a_2e_2 = b_1e_1 + b_2e_2.$$

Subtraindo a segunda expressão de v da primeira,

$$(a_1 - b_1)e_1 + (a_2 - b_2)e_2 = 0,$$

que equivale a

$$(a_1 - b_1)e_1 = -(a_2 - b_2)e_2.$$

Como e_1 e e_2 não são colineares, esta igualdade só é viável quando

$$a_1 - b_1 = a_2 - b_2 = 0,$$

porque o vetor nulo é o único que é múltiplo de qualquer outro. Portanto, as igualdades em (5) implicam que $a_1 = b_1$ e $a_2 = b_2$, mostrando que cada vetor corresponde a um único par de coordenadas em uma base ortonormal ε .

Vejamos de que forma as coordenadas se comportam relativamente à soma de vetores e ao produto de um vetor por um escalar. Sejam v_1 e v_2 dois vetores do plano cujas coordenadas relativamente a uma base ortonormal $\varepsilon = \{e_1, e_2\}$ são

$$v_1 = (a_1, b_1) \text{ e } v_2 = (a_2, b_2).$$

Pela definição de coordenadas, temos que

$$v_1 = a_1e_1 + b_1e_2 \text{ e } v_2 = a_2e_1 + b_2e_2.$$

Pela associatividades da adição de vetores

$$v_1 + v_2 = a_1e_1 + a_2e_1 + b_1e_2 + b_2e_2;$$

que, pelas propriedades do produto por escalar, podemos reescrever como

$$v_1 + v_2 = (a_1 + a_2)e_1 + (b_1 + b_2)e_2.$$

Logo, a $v_1 + v_2$ corresponde o par de coordenadas

$$(a_1 + a_2, b_1 + b_2).$$

Um argumento semelhante mostra que, se λ é um número real, então

$$\lambda \cdot v_1 = (\lambda a_1, \lambda b_1).$$

Expressaremos isto dizendo que a adição de vetores e a multiplicação de um vetor por um escalar são feitas “coordenada a coordenada”.

Agora que sabemos escrever vetores usando coordenadas, podemos expressar a equação de uma reta r em termos das coordenadas de um de seus vetores não nulos u . Assim, se $(x, y) \in r$ e (a, b) são as coordenadas de u relativamente à base ortonormal ε , então por (2),

$$(x, y) = t(a, b),$$

para algum número real t . Pela unicidade das coordenadas, isto equivale ao sistema

$$\begin{aligned}x &= at \\y &= bt.\end{aligned}$$

Como o número real t pode tomar qualquer valor que desejemos, sempre produzindo um ponto da reta para cada valor escolhido, dizemos que é um *parâmetro*. Por isso, as duas equações acima para x e y são chamadas de *equações paramétricas* da reta r .

1.3. Produto interno. Em física aprendemos que o *produto interno* (ou *produto escalar*) entre dois vetores v_1 e v_2 do plano é definido como sendo o número

$$(6) \quad \langle v_1 | v_2 \rangle = \|v_1\| \|v_2\| \cos \theta;$$

em que θ é o *menor* ângulo entre os vetores v_1 e v_2 . Se $\{e_1, e_2\}$ é uma base ortonormal do plano formada por vetores unitários perpendiculares entre si, de que maneira podemos expressar $\langle v_1 | v_2 \rangle$ em função das coordenadas de v_1 e v_2 relativas a esta base?

Para isto precisamos relacionar o ângulo θ aos ângulos que v_1 e v_2 formam com o vetor e_1 , e que são usados para determinar suas coordenadas. Chamando de α_1 e α_2 os ângulos entre e_1 e os vetores v_1 e v_2 , respectivamente, temos da figura que $\theta = \alpha_2 - \alpha_1$.

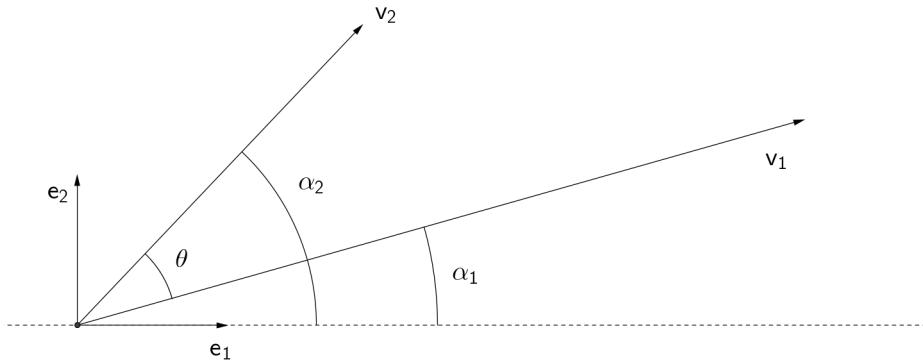


FIGURA 5. Produto interno

Portanto,

$$\cos(\theta) = \cos(\alpha_2 - \alpha_1) = \cos(\alpha_2) \cos(\alpha_1) + \sin(\alpha_2) \sin(\alpha_1),$$

de modo que

$$(7) \quad \langle v_1 | v_2 \rangle = \|v_1\| \cos(\alpha_1) \|v_2\| \cos(\alpha_2) + \|v_1\| \sin(\alpha_1) \|v_2\| \sin(\alpha_2).$$

Mas, se

$v_1 = (a_1, b_1) = (\|v_1\| \cos(\alpha_1), \|v_1\| \sin(\alpha_1))$ e $v_2 = (a_2, b_2) = (\|v_2\| \cos(\alpha_2), \|v_2\| \sin(\alpha_2))$, de modo que a equação (7) pode ser reescrita na forma

$$(8) \quad \langle v_1 | v_2 \rangle = a_1 a_2 + b_1 b_2.$$

Esta expressão do produto interno é muito conveniente. Por exemplo, a partir dela podemos provar facilmente as seguintes propriedades:

- (1) $\langle v_1 | v_2 + v_3 \rangle = \langle v_1 | v_2 \rangle + \langle v_1 | v_3 \rangle;$
- (2) $\langle v_1 | \lambda v_2 \rangle = \lambda \langle v_1 | v_2 \rangle;$
- (3) $\langle v_1 | v_2 \rangle = \langle v_2 | v_1 \rangle;$
- (4) $\langle v_1 | v_1 \rangle \geq 0;$
- (5) $\langle v_1 | v_1 \rangle = 0$ se, e somente se, $v_1 = 0$;

quaisquer que sejam os vetores v_1 , v_2 e v_3 do plano e o escalar λ . Note que a propriedade (3) implica que valem os análogos de (1) e (2) com a operação sendo efetuada na primeira coordenada e a segunda coordenada estando fixa.

Quando comparado ao resultado da adição de vetores e da multiplicação de um vetor por um escalar, o número gerado pelo produto interno parece carecer de sentido geométrico. Mas este não é, de forma alguma, o caso. Ao contrário, o produto interno codifica duas informações geométricas extremamente importantes: a norma de um vetor e o ângulo entre vetores. O caso mais óbvio é o da norma, uma vez que

$$\langle v | v \rangle = \|v\|^2$$

pois o cosseno de zero é 1. Se v tiver coordenadas (a_1, a_2) relativamente à base canônica, então, (8) nos dá

$$\langle v | v \rangle = a_1^2 + a_2^2;$$

onde

$$\|v\| = \sqrt{\langle v | v \rangle} = \sqrt{a_1^2 + a_2^2}$$

que é o que o teorema de Pitágoras nos faria esperar.

Por outro lado, se w é um segundo vetor do plano então, por (6),

$$\cos(\theta) = \frac{\langle v | w \rangle}{\|v\| \|w\|},$$

em que θ é o menor ângulo entre os dois vetores. Naturalmente isto só nos permite calcular θ porque a fórmula (8) nos dá uma maneira alternativa de determinar o produto interno, que independe da norma dos vetores e do ângulo entre eles. Supondo, por exemplo, que

$$v = (4, 3) \quad \text{e} \quad w = (1, 1)$$

são as representações de v e w em coordenadas relativamente à base canônica, temos que

$$\cos(\theta) = \frac{\langle (1, 1) | (4, 3) \rangle}{5\sqrt{2}} = \frac{7}{5\sqrt{2}},$$

de modo que

$$\theta = \arccos\left(\frac{7}{5\sqrt{2}}\right).$$

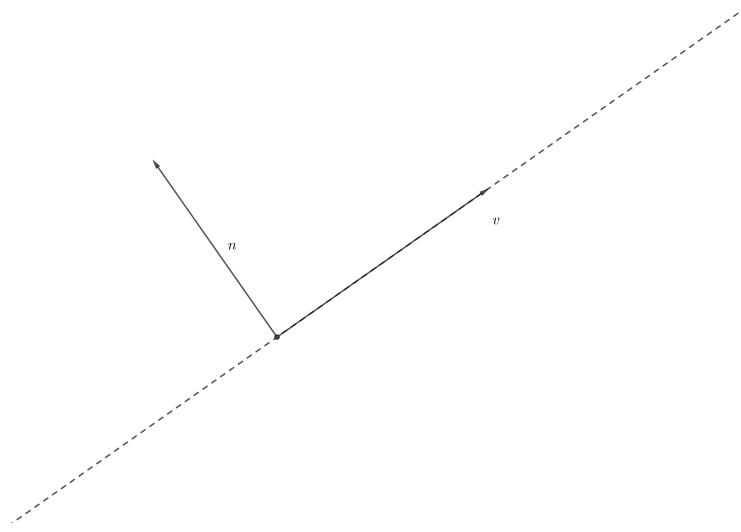


FIGURA 6. Reta construída a partir do vetor normal

Finalmente, a expressão (8) também nos permite interpretar geometricamente a equação cartesiana

$$\alpha x + \beta y = 0,$$

de uma reta pela origem, em que α e β são constantes. Se

$$n = (\alpha, \beta) \text{ e } v = (x, y),$$

então, por (8),

$$\langle n | v \rangle = \alpha x + \beta y.$$

Portanto, v pertence à reta de equação $\alpha x + \beta y = 0$ se, e somente se $\langle n | v \rangle = 0$. Em outras palavras,

v pertence à reta de equação $\alpha x + \beta y = 0$ se, e somente se v é perpendicular ao vetor fixo n .

Mas, por (2), podemos escrever qualquer vetor da reta r como múltiplo de um vetor fixo u , seu vetor diretor. Como $u \in r$, este vetor será necessariamente ortogonal a n . Portanto, se $n = (\alpha, \beta)$, como acima, então podemos tomar $u = (\beta, -\alpha)$ e vice-versa.

2. Operadores lineares

Até aqui podemos esticar ou encolher um vetor, multiplicando-o por um escalar, ou somar dois vetores; mas há muitas outras coisas que podemos fazer a um vetor, como rodá-lo ou refleti-lo relativamente a uma reta. O que não podemos fazer é entortá-lo, porque assim deixaria de ser um vetor. Além disso, como todos os vetores partem da origem, este ponto tem que ficar fixo por qualquer transformação de vetores. Nossa ponto de partida serão alguns exemplos cujo comportamento é fácil de descrever geometricamente.

2.1. Projeções. Começaremos pelas projeções, porque já vimos como calculá-las. Seja u um vetor *unitário* e v um vetor qualquer do plano. Por (3), a projeção de v em u é dada por

$$\text{Proj}_u(v) = (\|v\| \cos(\theta)) \cdot u,$$

em que θ é o ângulo entre u e v . Podemos usar o produto interno para reescrever esta fórmula como

$$(9) \quad \text{Proj}_u(v) = \langle u | v \rangle \cdot u,$$

uma vez que u tem módulo um. Disto obtemos, com muito pouco esforço, uma fórmula para a projeção em termos das coordenadas de u e de v . Se

$$u = (a, b) \text{ e } v = (x, y)$$

são as coordenadas destes vetores relativamente à base canônica, então

$$(10) \quad \text{Proj}_u(x, y) = ((ax + by)a, (ax + by)b) = (a^2x + aby, abx + b^2y).$$

Esta fórmula descreve uma aplicação do plano nele próprio que nos permite projetar sobre o vetor unitário u qualquer vetor do plano que desejarmos. Por exemplo, para escrever a fórmula da projeção sobre o vetor $(1, 1)$, começamos por normalizá-lo, obtendo

$$u = \frac{(1, 1)}{\sqrt{2}}.$$

Substituindo $1/\sqrt{2}$ em a e b , obtemos

$$\text{Proj}_u(x, y) = \frac{1}{2}(x + y, x + y);$$

de modo que

$$\text{Proj}_u(7, 11) = (9, 9) \quad \text{e} \quad \text{Proj}_u(2, 3) = \frac{5}{2}(1, 1).$$

A fórmula da projeção a partir do produto interno nos dá uma maneira bastante simples de calcular as coordenadas de um vetor v relativamente a uma base ortonormal

$$\beta = \{u_1, u_2\}$$

quando as coordenadas de v , u_1 e u_2 relativamente à base canônica são conhecidas. Lembre-se que, por definição, a_1 e a_2 são as coordenadas de v relativamente a β se

$$a_1 u_1 = \text{Proj}_{u_1}(v) \quad \text{e} \quad a_2 u_2 = \text{Proj}_{u_2}(v).$$

Podemos, então, deduzir de (9) que

$$a_1 = \langle u_1 | v \rangle \quad \text{e} \quad a_2 = \langle u_2 | v \rangle.$$

Por exemplo, para calcular as coordenadas de $v = (1, 2)$ relativamente à base ortonormal

$$\beta = \left\{ \frac{1}{5}(4, 3), \frac{1}{5}(3, -4) \right\},$$

calculamos

$$\frac{1}{5}\langle(1, 2) | (4, 3)\rangle = 2 \quad \text{e} \quad \frac{1}{5}\langle(1, 2) | (3, -4)\rangle = -1$$

de modo que $v = 2u_1 - u_2$.

2.2. Reflexões. As reflexões podem ser tratadas de maneira semelhante às projeções. Chamaremos de *espelho* a reta em torno do qual será realizada a reflexão e cujo vetor diretor *unitário* denotaremos por u . Já n será um vetor, também unitário, perpendicular a u . No plano, uma vez fixado u , só há duas possibilidades para n , porque a norma de n está fixa, pois é igual a um, e sua direção também, já que está sobre a reta perpendicular ao espelho. Resta escolher seu sentido, para o qual temos apenas duas possibilidades. Observe que os vetores u e n formam uma base ortonormal do plano, de acordo com a definição do artigo 1.2, pois são unitários e perpendiculares entre si. Portanto, se v for um vetor do plano e $R(v)$ seu reflexo relativamente ao espelho de vetor diretor u , temos que

$$R(v) = \text{Proj}_u(R(v)) + \text{Proj}_n(R(v)).$$

Resta-nos determinar as projeções de $R(v)$ sobre u e n em termos das coordenadas de v nesta base. Para isto faremos uso da descrição geométrica usual de uma reflexão.

Para começar, um vetor v e seu reflexo $R(v)$ têm ambos o mesmo módulo. Além disso, o ângulo que o vetor v forma com o espelho é o mesmo entre $R(v)$ e o espelho. A diferença é que v está de um lado do espelho, ao passo que $R(v)$ está do outro lado, como ilustra a figura 7.

O ponto crucial para determinar uma fórmula para o reflexo $R(v)$ de um vetor v é observar que as projeções de v e $R(v)$ sobre u satisfazem

$$\text{Proj}_u(R(v)) = \text{Proj}_u(v).$$

ao passo que as projeções sobre a normal n satisfazem,

$$\text{Proj}_n(R(v)) = -\text{Proj}_n(v).$$

Logo,

$$R(v) = \text{Proj}_u(v) - \text{Proj}_n(v) = v - 2\text{Proj}_n(v),$$

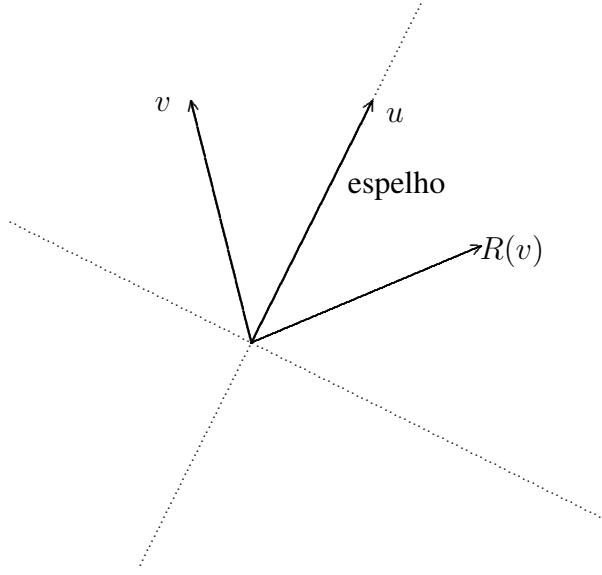


FIGURA 7. Reflexão

pois

$$v = \text{Proj}_u(v) + \text{Proj}_n(v).$$

Representando a projeção sobre n em termos do produto interno, como no artigo anterior, temos que

$$\text{Proj}_n(v) = \langle n | v \rangle \cdot n,$$

onde

$$(11) \quad R(v) = v - 2\langle n | v \rangle \cdot n.$$

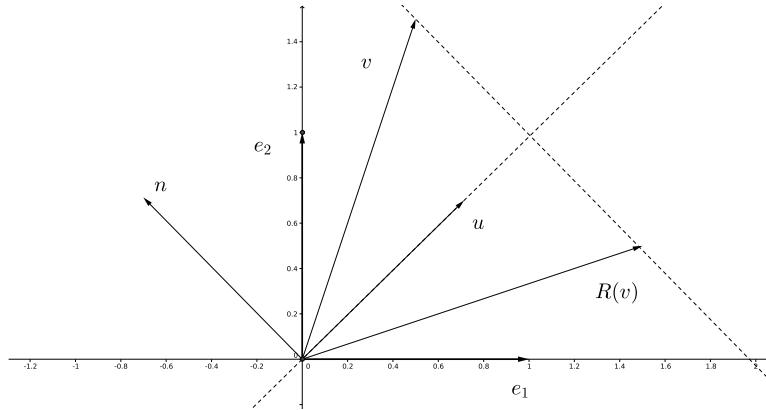
Encerraremos este artigo determinando uma fórmula para $R(v)$ em função das coordenadas de v . A maneira mais fácil de fazer isto consiste em usar a base ortonormal $\beta = \{u, n\}$ ao descrever as coordenadas dos vetores. Afinal, relativamente a esta base, u tem coordenadas $(1, 0)$ e n tem coordenadas $(0, 1)$, pois

$$u = 1 \cdot u + 0 \cdot n \quad \text{e} \quad n = 0 \cdot u + 1 \cdot n.$$

Supondo que $(v)_\beta = (x, y)$, uma aplicação direta da fórmula (11) nos dá

$$(R(x, y))_\beta = (x, y) - 2y(0, 1) = (x, -y);$$

como seria de esperar da descrição geométrica. O problema é que ao usar $\beta = \{u, n\}$ como base ortonormal estamos criando uma situação um pouco artificial. Na prática, os vetores u e v são dados em termos de suas coordenadas relativamente à base canônica, e não vice-versa. Portanto, tendo em vista futuras aplicações, convém determinar como seria a fórmula da reflexão em termos das coordenadas dos vetores relativamente a uma base ortonormal qualquer.

FIGURA 8. Reflexão em torno da reta $y = x$.

Para isto suporemos que o vetor unitário u tem coordenadas (a, b) relativamente à base canônica. Mas, o produto interno de u com $n = (-b, a)$ é igual a zero e

$$\|n\| = \sqrt{a^2 + b^2} = \|u\| = 1,$$

de modo que n é um vetor unitário ortogonal a u . Portanto, se $v = (x, y)$, a fórmula da reflexão obtida acima nos diz que

$$R(x, y) = (x, y) - 2(ay - bx) \cdot (-b, a);$$

isto é,

$$(12) \quad R(x, y) = ((1 - 2b^2)x + 2aby, 2abx + (1 - 2a^2)y).$$

Por exemplo, se

$$u = \frac{1}{\sqrt{2}}(1, 1), \quad \text{então}, \quad n = \frac{1}{\sqrt{2}}(-1, 1)$$

e a fórmula da reflexão em torno da reta $y = x$ cujo vetor diretor é u será $R(x, y) = (y, x)$, como ilustrado na figura 8.

2.3. Rotação. Passando à rotação, digamos que ρ_θ seja a função que roda um vetor v de um ângulo θ no sentido anti-horário. Mais uma vez, nosso objetivo consiste em escrever uma fórmula para este operador em termos das coordenadas de um vetor v relativamente à base canônica. Se $v = (x, y)$, então a fórmula (4) nos permite afirmar que

$$x = \|v\| \cos \alpha \quad \text{e} \quad y = \|v\| \sin \alpha$$

em que α é o ângulo entre v e o vetor e_1 . Tendo expresso x e y desta maneira, fica fácil determinar as coordenadas de $\rho_\theta(v)$. Afinal, ao rodar v de um ângulo θ no sentido anti-horário, o ângulo entre v e e_1 aumenta de α para $\alpha + \theta$. Isto é, as coordenadas de $\rho_\theta(v)$ serão

$$(\|v\| \cos(\alpha + \theta), \|v\| \sin(\alpha + \theta)).$$

Para explicitar a relação entre estas coordenadas e as coordenadas x e y de v , usamos duas bem conhecidas fórmulas de trigonometria

$$\begin{aligned}\operatorname{sen}(\alpha + \theta) &= \operatorname{sen}(\alpha)\cos(\theta) + \operatorname{sen}(\theta)\cos(\alpha) \quad \text{e} \\ \cos(\alpha + \theta) &= \cos(\alpha)\cos(\theta) - \operatorname{sen}(\theta)\sin(\alpha).\end{aligned}$$

Multiplicando estas expressões por $\|v\|$ e substituindo $\|v\|\cos\alpha$ por x e $\|v\|\sin\alpha$ por y , obtemos

$$\rho_\theta(x, y) = (\cos(\theta)x - \operatorname{sen}(\theta)y, \operatorname{sen}(\theta)x + \cos(\theta)y),$$

que é a fórmula desejada.

2.4. Cisalhamento. Nossa último exemplo são os cisalhamentos. O fenômeno mais conhecido, em que um cisalhamento se manifesta de maneira visível, ocorre quando nuvens são dispersadas pelo vento que sopra com diferentes velocidades ao longo de uma certa direção.



FIGURA 9. Cisalhamento do vento (Fonte: CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=121109>).

Na figura 9, os cristais microscópicos de gelo de uma nuvem do tipo cirrus são arrastados por um cisalhamento de vento formando nuvens em forma de plumas. O cisalhamento do vento também é conhecido como cortante do vento, ou pela expressão inglesa “wind shear”. Para obter uma ideia mais clara do efeito de um cisalhamento, imagine que o vento sopra paralelamente ao eixo horizontal, mas que sua velocidade aumenta à medida que nos afastamos deste eixo. Na figura 10, representamos a nuvem de cirrus como um segmento de reta vertical e supomos que o vento sopra horizontalmente, da esquerda para a direita, como ilustrado pelas semiretas tracejadas. Como o diagrama sugere, ao longo do tempo cada partícula da nuvem é arrastada para mais longe dela pelo efeito do vento.

Dado um número real α , o operador linear de cisalhamento c_α , definido pela regra

$$c_\alpha(x, y) = (x + \alpha y, y),$$

tenta reproduzir o efeito do cisalhamento do vento em uma nuvem. Para mostrar que c_α tem o efeito desejado, basta aplicá-lo a um múltiplo do vetor e_2 , porque isto nos dá

$$c_\alpha(0, y) = (\alpha y, y) = y(\alpha, 1),$$

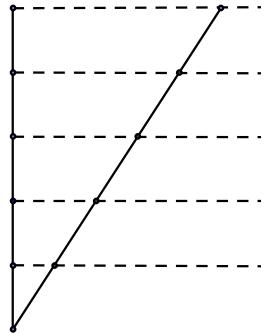


FIGURA 10. Diagrama de cisalhamento pelo vento.

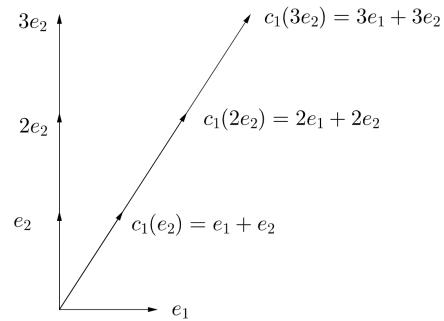


FIGURA 11. Efeito de c_1 em e_2 , $2e_2$ e $3e_2$.

de modo que o vetor é tanto mais alongado ao longo da horizontal, quanto mais longe estiver da origem. O efeito de c_1 sobre alguns vetores verticais é ilustrado na figura 11.

2.5. Operadores lineares. Uma coisa que transparece das fórmulas obtidas para projeções, reflexões e rotações é que as coordenadas são sempre expressões lineares sem termo constante, nas coordenadas x e y do argumento v . Os operadores com esta propriedade são tão abundantes nas ciências naturais, e tão importantes no estudo dos vetores, que merecem uma designação à parte.

Seja T uma função do plano nele mesmo e fixemos uma base ortonormal do plano. Diremos que T é um *operador linear do plano* se existirem constantes a , b , c e d de modo que a imagem de qualquer vetor v pode ser escrita na forma

$$T(v) = (ax + by, cx + dy) \text{ sempre que } v = (x, y).$$

A origem do uso do adjetivo linear para designar tais operadores é claro: as coordenadas do vetor imagem são, de fato, expressões lineares em x e y . Observe que excluímos a possibilidade de termos constantes nesta expressão desde o começo, porque decidimos de partida que a imagem do vetor zero por T teria que ser o mesmo vetor zero, já que todos os vetores partem de um mesmo ponto.

Os operadores lineares do plano têm três propriedades importantes. Se v_1 e v_2 são dois vetores quaisquer do plano e $\lambda \in \mathbb{R}$, então:

- (1) $T(0) = 0$;
- (2) $T(v_1 + v_2) = T(v_1) + T(v_2)$;
- (3) $T(\lambda v_1) = \lambda T(v_1)$.

A propriedade (1) é óbvia; provaremos a segunda e deixaremos a terceira aos seus cuidados. Suponhamos que v_1 e v_2 têm coordenadas

$$v_1 = (x_1, y_1) \text{ e } v_2 = (x_2, y_2),$$

relativamente à base ortonormal fixada. Neste caso,

$$v_1 + v_2 = (x_1 + x_2, y_1 + y_2);$$

de modo que

$$T(v_1 + v_2) = (a(x_1 + x_2) + b(y_1 + y_2), c(x_1 + x_2) + d(y_1 + y_2)).$$

Mas o lado direito da equação acima é

$$((ax_1 + by_1) + (ax_2 + by_2), (cx_1 + dy_1) + (cx_2 + dy_2))$$

que é igual à soma de vetores,

$$(ax_1 + by_1, cx_1 + dy_1) + (ax_2 + by_2, cx_2 + dy_2);$$

isto é, a $T(v_1) + T(v_2)$, provando assim a propriedade desejada.

Na verdade, qualquer função do plano nele mesmo que satisfaz estas três propriedades tem que ser um operador linear. A verificação é simples e muito importante para a caracterização final que daremos a estes operadores, por isso vamos fazê-la em detalhes. Para deixar bem claro o que queremos fazer, convém enunciá-lo de maneira bastante precisa:

se um operador T do plano satisfaz as propriedades (1), (2) e (3) acima então existem constantes a, b, c e d de modo que $T(x, y) = (ax + by, cx + dy)$.

Começaremos supondo que $v = (x, y)$ é um vetor escrito na base canônica $\varepsilon = \{e_1, e_2\}$. Por definição, isto significa que

$$v = xe_1 + ye_2.$$

Portanto,

$$T(v) = T(xe_1 + ye_2).$$

Usando as propriedades (2) e (3) o lado direito desta última equação pode ser escrito na forma

$$T(v) = xT(e_1) + yT(e_2).$$

Supondo que

$$T(e_1) = ae_1 + ce_2 \text{ e que } T(e_2) = be_1 + de_2,$$

obtemos

$$T(v) = x(ae_1 + ce_2) + y(be_1 + de_2) = (ax + by)e_1 + (cx + dy)e_2.$$

que equivale a

$$T(x, y) = (ax + by, cx + dy).$$

Observe que, como os vetores e_1 e e_2 estão fixos, os números reais a , b , c e d dependem apenas de T e não das coordenadas de v . Na verdade, descobrimos o que estes quatro números representam: são as coordenadas de $T(e_1)$ e $T(e_2)$. Voltaremos a usar isto em 3.1. Convém resumir o que fizemos acima para uso futuro.

PROPOSIÇÃO 2.1. *Se T é uma função do plano no plano então as seguintes condições são equivalentes:*

- T satisfaz as propriedades (1), (2) e (3) acima;
- $T(x, y) = (ax + by, cx + dy)$ em que $T(e_1) = (a, c)$ e $T(e_2) = (b, d)$.

2.6. Combinando operadores lineares. Dois operadores lineares do plano podem ser somados ou compostos, disto resultando um novo operador linear do plano. Se S e T são operadores lineares do plano, então definimos sua soma como sendo a função $S + T$ definida em um vetor v por

$$(S + T)(v) = S(v) + T(v),$$

ao passo que sua composição $S \circ T$ é definida por

$$(S \circ T)(v) = S(T(v)).$$

Note que $S \circ T$ e $T \circ S$ representam operadores que podem ser diferentes, ao passo que $S + T$ e $T + S$ sempre designam o mesmo operador, porque a soma de vetores é comutativa.

Verificaremos com cuidado que tanto $S + T$ quanto $S \circ T$ são operadores lineares. A maneira mais fácil seria provar que estas operadores satisfazem as propriedades (1), (2) e (3) do artigo anterior. Pela proposição 2.1 isto garantiria que se tratam de operadores lineares. Em vez disso, vamos deduzir uma fórmula em termos de coordenadas para $S + T$ e $S \circ T$, a partir das respectivas fórmulas para S e T . Procederemos assim porque estas fórmulas serão necessárias no artigo seguinte.

Supondo fixada uma base ortonormal do plano, digamos que

$$T(x, y) = (ax + by, cx + dy) \text{ e que } S(x, y) = (\alpha x + \beta y, \gamma x + \delta y),$$

em que $a, b, c, d, \alpha, \beta, \gamma, \delta$ são constantes. Por definição,

$$(S + T)(x, y) = S(x, y) + T(x, y),$$

que é igual a

$$(ax + by, cx + dy) + (\alpha x + \beta y, \gamma x + \delta y);$$

somando os vetores, concluímos que

$$(S + T)(x, y) = ((a + \alpha)x + (b + \beta)y, (c + \gamma)x + (d + \delta)y).$$

A fórmula resultante é muito fácil de lembrar, porque apenas somamos os coeficientes de x e y em cada coordenada do vetor imagem. A fórmula da composta, infelizmente, está longe de ser tão simples. Partindo da definição temos que

$$(S \circ T)(x, y) = S(T(x, y)) = S(ax + by, cx + dy).$$

Aplicando a fórmula para S ,

$$(S \circ T)(x, y) = (\alpha(ax + by) + \beta(cx + dy), \gamma(ax + by) + \delta(cx + dy)).$$

Reagrupando os termos,

$$(13) \quad (S \circ T)(x, y) = ((\alpha a + \beta c)x + (\alpha b + \beta d)y, (\gamma a + \delta c)x + (\gamma b + \delta d)y),$$

que é bem menos fácil de lembrar do que a anterior.

3. Matriz de um operador linear

A fórmula (13), do final da seção anterior, nos permite encontrar a composta de dois operadores lineares. Embora não seja difícil, esta fórmula não é fácil de lembrar, sobretudo quando precisamos compor não dois, mas três ou quatro operadores. Problemas que levam a este tipo de cálculo ocorrem frequentemente na prática e levaram Arthur Cayley, no século XIX, a procurar uma maneira mais prática de resolvê-los. Para isto ele inventou as matrizes.

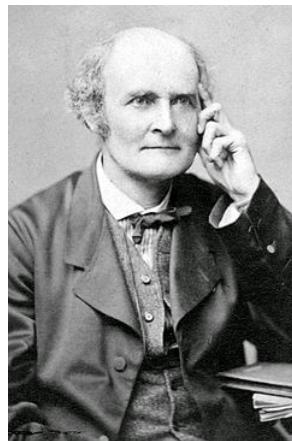


FIGURA 12. Arthur Cayley



FIGURA 13. J. J. Sylvester

3.1. Matriz de um operador linear. A ideia de Cayley é que, uma vez fixada uma base ortonormal do plano, um operador linear fica completamente determinado por quatro números: os coeficientes de x e y nas expressões que definem as coordenadas de $T(x, y)$. Quando

$$T(x, y) = (a_{11}x + a_{12}y, a_{21}x + a_{22}y),$$

os números são a_{11} , a_{12} , a_{21} e a_{22} . Mas isto significa que, para fazer cálculos com T basta conhecer estes números e descobrir como se transformam sob estes cálculos. Para tornar tudo mais transparente, Cayley resolveu dispor estes números em um quadro,

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

e assim foram inventadas as matrizes. Como sempre, este resumo histórico não representa o que realmente aconteceu. A disposição em forma de quadro já era usada desde o século XVIII para denotar determinantes, e o nome matriz foi usado por Sylvester antes mesmo do primeiro artigo do Cayley sobre o assunto. Mais detalhes podem ser encontrados em [2, p. 171].

Como a matriz de um operador linear depende completamente da base ortonormal do plano que foi escolhida e fixada, denotaremos a matriz de T escrita acima por $(T)_\varepsilon$, em que $\varepsilon = \{e_1, e_2\}$ é a base canônica. Como,

$$T(e_1) = T(1, 0) = (a_{11}, a_{21}) \quad \text{e} \quad T(e_2) = T(0, 1) = (a_{12}, a_{22}),$$

vemos que as colunas de

$$(14) \quad (T)_\varepsilon = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

correspondem à imagem de e_1 e e_2 por T . Podemos ilustrar isto escrevendo

$$(15) \quad (T)_\varepsilon = \begin{bmatrix} | & | \\ (Te_1)_\varepsilon & (Te_2)_\varepsilon \\ | & | \end{bmatrix}.$$

Quando a fórmula de um operador linear é conhecida, a melhor maneira de construir sua matriz é ler as entradas diretamente da fórmula, como fizemos com o operador T no início deste artigo. Por exemplo, usando as fórmulas (10) e (12) obtidas para a projeção e reflexão, vemos que suas matrizes são

$$(\text{Proj}_u)_\varepsilon = \begin{bmatrix} a^2 & ab \\ ab & b^2 \end{bmatrix} \quad \text{e} \quad (R)_\varepsilon = \begin{bmatrix} (1 - 2b^2) & 2ab \\ 2ab & (1 - 2a^2) \end{bmatrix}.$$

O problema com estas fórmulas é que não são exatamente fáceis de lembrar. Na verdade, quando o operador admite uma descrição geométrica simples, é frequentemente mais fácil

construir sua matriz a partir da ação do operador sobre os vetores de uma base. Por exemplo, um cisalhamento horizontal não afeta os vetores ao longo da horizontal, ao passo que os vetores verticais são inclinados. Portanto, usando a notação do artigo 2.4,

$$c_\alpha(e_1) = e_1 \quad \text{e} \quad c_\alpha(e_2) = e_2 + \alpha e_1$$

Escrevendo estes vetores na base canônica, temos

$$c_\alpha(1, 0) = (1, 0) \quad \text{e} \quad c_\alpha(0, 1) = (\alpha, 1).$$

Portanto, a matriz deste cisalhamento na base canônica será

$$(c_\alpha)_\varepsilon = \begin{bmatrix} 1 & \alpha \\ 0 & 1 \end{bmatrix}$$

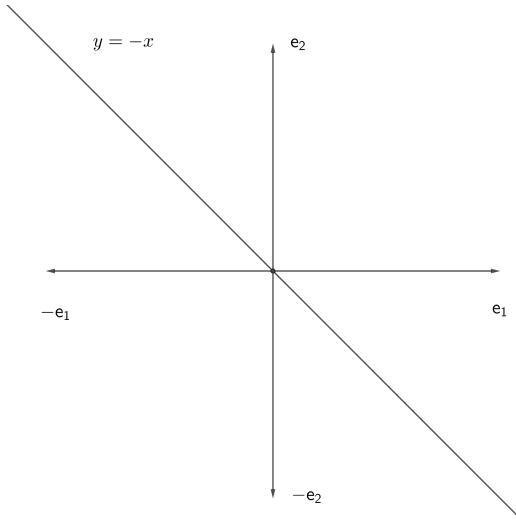


FIGURA 14. Reflexão em torno de $y = -x$.

Outro exemplo simples é a reflexão R cujo espelho é a reta ℓ de equação $y = -x$. Como ilustrado na figura 14, o ângulo entre e_1 e ℓ é igual ao ângulo entre $-e_2$ e ℓ . Já o ângulo entre e_2 e ℓ é igual ao ângulo entre $-e_1$ e ℓ . Logo,

$$R(e_1) = -e_2 \quad \text{e} \quad R(e_2) = -e_1.$$

Portanto, a matriz de R relativamente à base canônica será

$$(R)_\varepsilon = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$$

Na verdade, no caso da projeção e da reflexão, a matriz pode ser expressa de maneira ainda mais compacta usando as operações com matrizes, como veremos no artigo 3.2.

Outro operador cuja matriz é facilmente construída diretamente da descrição geométrica é a rotação. Como mostra a figura 15, as coordenadas do vetor $\rho_\theta(e_1)$ são dadas pelos catetos do triângulo retângulo AOB . Como as rotações preservam o comprimento dos vetores, a hipotenusa \overline{OA} tem comprimento igual a

$$\|\rho_\theta(e_1)\| = \|e_1\| = 1,$$

de modo que \overline{OB} tem comprimento $\cos(\theta)$ e \overline{AB} tem comprimento $\sin(\theta)$; o que nos dá

$$\rho_\theta(e_1) = (\cos(\theta), \sin(\theta))$$

Como o triângulo ODC é congruente a AOB , os catetos \overline{OC} e \overline{DC} terão comprimen-

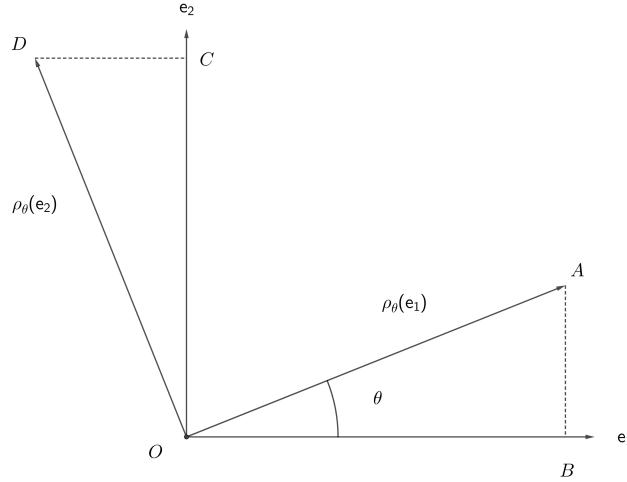


FIGURA 15. Rotação anti-horária de ângulo θ .

tos iguais a $\cos(\theta)$ e $\sin(\theta)$, respectivamente. Entretanto, o segmento \overline{DC} está do lado negativo do eixo horizontal, de modo que a primeira entrada de $\rho_\theta(e_2)$ deve ser negativa. Levando isto em conta, temos que

$$\rho_\theta(e_2) = (-\sin(\theta), \cos(\theta)).$$

Pondo estes vetores nas colunas da matriz, encontramos

$$(16) \quad (\rho_\theta)_\varepsilon = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

Vale a pena comparar este argumento com o que fizemos no artigo 2.3 para achar diretamente a fórmula da rotação. Lá, tivemos que usar as fórmulas para o seno e o cosseno de uma soma de ângulos; aqui, nada além das definições de seno e cosseno.

Para encerrar, descreveremos explicitamente as matrizes correspondentes à soma e à composição de dois operadores cujas matrizes são conhecidas. Para isto, considere dois

operadores lineares T e S do plano, definidas em uma base ε por

$$T(x, y) = (ax + by, cx + dy) \text{ e } S(x, y) = (\alpha x + \beta y, \gamma x + \delta y).$$

Pela regra criada por Cayley as matrizes correspondentes na base ε serão

$$(T)_\varepsilon = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ e } (S)_\varepsilon = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}.$$

Usando as fórmulas para $S + T$ e $S \circ T$ obtidas no artigo anterior, temos que

$$(S + T)_\varepsilon = \begin{bmatrix} a + \alpha & b + \beta \\ c + \gamma & d + \delta \end{bmatrix} \text{ ao passo que } (S \circ T)_\varepsilon = \begin{bmatrix} \alpha a + \beta c & \alpha b + \beta d \\ \gamma a + \delta c & \gamma b + \delta d \end{bmatrix}.$$

Cayley deu, então, um passo à frente, utilizando estas fórmulas para definir a adição e a multiplicação das próprias matrizes, de modo que

$$(T)_\varepsilon + (S)_\varepsilon = (S + T)_\varepsilon \text{ e } (S)_\varepsilon \cdot (T)_\varepsilon = (S \circ T)_\varepsilon.$$

Abstraindo completamente dos operadores, obtemos desta forma as operações usuais com matrizes: a soma, definida entrada a entrada, e a multiplicação, definida pela regra

$$\begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \cdot \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} \alpha a + \beta c & \alpha b + \beta d \\ \gamma a + \delta c & \gamma b + \delta d \end{bmatrix}.$$

Portanto, a regra para multiplicação de matrizes, à primeira vista tão artificial, é obtida coletando os coeficientes de x e y na fórmula resultante da composição de dois operadores lineares. Antes de passarmos adiante, é importante lembrar que a composição de operadores não é comutativa. Como a multiplicação de matrizes apenas põe em uma tabela os coeficientes da composição, ela também não será comutativa. De fato,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} = \begin{bmatrix} a\alpha + b\gamma & a\beta + b\delta \\ c\alpha + d\gamma & c\beta + d\delta \end{bmatrix} \neq \begin{bmatrix} \alpha a + \beta c & \alpha b + \beta d \\ \gamma a + \delta c & \gamma b + \delta d \end{bmatrix}.$$

3.2. Matrizes retangulares. Ainda que você tenha se convencido de que as matrizes quadradas 2×2 possam ser úteis na representação de operadores lineares do plano, talvez você esteja se perguntando se matrizes retangulares não são fruto da obsessão dos matemáticos em generalizar tudo o que pode ser generalizado. Apesar de ter todo o resto deste livro para lhe convencer de que não é este o caso, não custa dar alguns exemplos relacionados aos vetores do plano e seus operadores lineares.

Para começar, podemos considerar um vetor do plano como sendo uma matriz. À primeira vista o natural seria descrever um vetor como sendo uma matriz 1×2 , mas a verdade é que é melhor identificar um vetor (a, b) com a matriz

$$\begin{bmatrix} a \\ b \end{bmatrix},$$

de tamanho 2×1 , também conhecida como *matriz coluna*. A razão para esta escolha um tanto bizarra logo ficará clara. O fato é que, somando vetores, ou as matrizes que lhes correspondem, obtemos o mesmo resultado. Mais precisamente,

a matriz coluna correspondente à soma dos vetores u com v é igual à soma das matrizes coluna correspondentes a u e v , e o mesmo pode ser dito sobre o produto de um vetor por um escalar.

Por isso, de agora em diante, consideraremos vetores do plano como sendo matrizes 2×1 sempre que isto for conveniente. Supondo que é este o caso para os vetores u e v , podemos descrever seu produto escalar, a partir do produto de matrizes, por

$$(17) \quad \langle u | v \rangle = u^t \cdot v.$$

Note que convertemos o vetor coluna u em um vetor linha tomado a sua transposta, para que fosse possível efetuar a multiplicação desejada.

Passando aos operadores lineares do plano, vimos que se T é definida, na base canônica, por

$$T(x, y) = (ax + by, cx + dy)$$

então a matriz a ela associada é

$$(T)_\varepsilon = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

No entanto, um cálculo simples mostra que se o vetor v tem coordenadas

$$v = \begin{bmatrix} x \\ y \end{bmatrix},$$

na mesma base ε , então as coordenadas de Tv nesta mesma base serão

$$Tv = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

 Com isto podemos explicar porque escolhemos representar vetores como colunas e não linhas. Lembre-se que a um operador linear do plano fizemos corresponder uma matriz 2×2 . Vetores escritos como matrizes linha têm tamanho 1×2 o que nos obrigaria a multiplicá-los à esquerda das matrizes que designam os operadores. Mas isto produz um conflito com a convenção de que o argumento de um operador sempre aparece à direita do símbolo que a denota; assim Tv , e não vT . Para evitar a confusão que resultaria da permutação dos lados entre duas fórmulas que representam exatamente o mesmo fato, preferimos escrever os vetores como colunas, em vez de linhas.

Mais interessante ainda são as expressões para as matrizes das projeções e reflexões que obtemos combinando a multiplicação e a transposição. Por exemplo, vimos no artigo

2.1 que a projeção de um vetor v qualquer sobre um vetor unitário u é dada por

$$\text{Proj}_u(v) = \langle u | v \rangle \cdot u,$$

que (17) nos permite reescrever na forma

$$\text{Proj}_u(v) = (u^t \cdot v) \cdot u.$$

Como $u^t \cdot v$ é um escalar, esta fórmula é igual a

$$\text{Proj}_u(v) = u \cdot (u^t \cdot v).$$

Donde, pelas propriedades do produto de matrizes, obtemos

$$\text{Proj}_u(v) = (u \cdot u^t) \cdot v.$$

Portanto, a matriz que descreve a projeção de v em u é igual ao produto $u \cdot u^t$. De fato, supondo que $u = [a, b]^t$ e efetuando o produto, chegamos à mesma matriz que havíamos obtido na equação (17).

Por outro lado, como vimos no artigo 2.2, a reflexão do vetor v relativamente à reta pela origem de vetor unitário normal n é igual a

$$R(v) = v - 2 \text{Proj}_n(v).$$

Aplicando a fórmula matricial obtida acima a $\text{Proj}_n(v)$, obtemos

$$R(v) = v - 2nn^t v.$$

A presença do v em ambas as parcelas sugere pô-lo em evidência. No entanto, $(1 - 2nn^t)v$ não faz sentido, porque duas matrizes só podem ser somadas se têm o mesmo tamanho. Contudo, mesmo considerando o escalar 1 como uma matriz 1×1 , não podemos somá-lo à matriz uu^t , que tem tamanho 2×2 . Felizmente há uma saída simples, basta considerar v como sendo o produto $I \cdot v$, em que I é a matriz identidade 2×2 . Fazendo isto, obtemos

$$(18) \quad R(v) = (I - 2nn^t)v.$$

Mais uma vez, se n tem coordenadas $(b, -a)$ na base ε , um cálculo elementar mostra que $I - 2nn^t$ coincide com a matriz de reflexão encontrada no artigo 2.2. Obtemos, assim, fórmulas muito compactas para a projeção e reflexão no plano usando a multiplicação de matrizes não quadradas.

Encerraremos calculando a matriz do operador linear T do plano que consiste de uma rotação anti-horária de $\pi/3$ radianos seguida de uma reflexão R que tem a reta $y = 2x$ como espelho. Para isso, calcularemos, separadamente, as matrizes da rotação e da reflexão e multiplicaremos as duas na ordem indicada. Para obter a matriz da rotação basta usar a fórmula (16), que nos dá

$$(\rho_{\pi/3})_\varepsilon = \begin{bmatrix} \cos(\pi/3) & -\sin(\pi/3) \\ \sin(\pi/3) & \cos(\pi/3) \end{bmatrix} = \begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix}.$$

A reflexão R tem por espelho a reta de equação cartesiana $2x - y = 0$. Logo, pelo que vimos ao final do artigo 1.3,

$$n = \frac{1}{\sqrt{5}}(2, -1),$$

é um vetor unitário normal a esta reta. Portanto, pela fórmula (18),

$$(R)_\varepsilon = I - 2nn^t = \frac{1}{5} \begin{bmatrix} -3 & 4 \\ 4 & 3 \end{bmatrix}.$$

Como T foi definido como sendo o operador que consiste em aplicar a reflexão em seguida à rotação, temos que

$$T = R \circ \rho_\varepsilon$$

Mas já sabemos que

$$(R \circ \rho_{\pi/3})_\varepsilon = (R)_\varepsilon \cdot (\rho_{\pi/3})_\varepsilon,$$

de modo que

$$(T)_\varepsilon = \frac{1}{5} \begin{bmatrix} -3 & 4 \\ 4 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix} = \frac{1}{10} \begin{bmatrix} 4\sqrt{3}-3 & 3\sqrt{3}+4 \\ 3\sqrt{3}+4 & -4\sqrt{3}+3 \end{bmatrix}.$$

Esta é uma isometria porque uma rotação e uma reflexão são isometrias e a composta de quaisquer duas isometrias também é uma isometria. Além do mais, como o determinante desta matriz é igual a -1 , então T tem que ser uma reflexão.

3.3. Operadores inversíveis e operadores ortogonais. Muitos dos operadores que estudaremos podem ser desfeitos; isto é, admitem um operador inverso. Sejam S e T operadores do plano. Dizemos que S é o inverso de T se

$$S \circ T = T \circ S = \text{id},$$

em que id é o operador identidade do plano, definido pela regra $\text{id}(v) = v$. A fórmula acima pode ser reescrita na forma

$$(S)_\varepsilon(T)_\varepsilon = (T)_\varepsilon(S)_\varepsilon = I,$$

em que I é a matriz identidade 2×2 . Em outras palavras,

o operador S é o inverso do operador T se, e somente se, a matriz de S é a inversa da matriz de T ;

desde que ambas as matrizes estejam expressas relativamente a uma mesma base ortonormal do plano. Sempre que for conveniente denotaremos a inversa de um operador T por T^{-1} . Você encontrará uma fórmula para o cálculo da inversa de uma matriz 2×2 no exercício 17.

Muitos dos operadores que estudamos anteriormente admitem inversos. Este é o caso, por exemplo, das reflexões e das rotações. Das propriedades geométricas da reflexão vemos

que deve ser sua própria inversa, ao passo que a inversa de uma rotação anti-horária de ângulo θ é a rotação horária de mesmo ângulo. Denotando por R a reflexão e por ρ_θ a rotação, temos que

$$R^{-1} = R \text{ e que } \rho_\theta^{-1} = \rho_{(-\theta)} = \rho_{2\pi-\theta}.$$

Em ambos os casos a inversa da matriz do operador coincide com a sua transposta, pois $(R)_\varepsilon^t = (R)_\varepsilon$ e

$$(\rho_\theta)_\varepsilon^t = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}^t = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} = (\rho_{(-\theta)})_\varepsilon = (\rho_\theta)_\varepsilon^{-1},$$

já que o cosseno é uma função par, ao passo que o seno é ímpar. Dizemos que uma matriz é *ortogonal* se é inversível, e sua inversa coincide com sua transposta; um operador do plano cuja matriz é ortogonal também é chamado de ortogonal.

Um exemplo que ilustra bem a utilidade de inverter operadores ocorre quando desejamos calcular, por exemplo, um operador linear S do plano que leva o retângulo de vértices

$$(0, 0), (1, 2), (-2, 1) \quad \text{e} \quad (-1, 3)$$

no quadrado de vértices

$$(0, 0), (1, 0), (0, 1) \quad \text{e} \quad (1, 1).$$

Note, em primeiro lugar, que, como o operador é linear, se

$$(19) \quad S(1, 2) = (1, 0) \quad \text{e} \quad S(-2, 1) = (0, 1)$$

então, obrigatoriamente,

$$S(-1, 3) = S((1, 2) + (-2, 1)) = S(1, 2) + S(-2, 1) = (1, 1).$$

Contudo, pela definição do inverso de um operador, temos de (19) que

$$S^{-1}(1, 0) = (1, 2) \quad \text{e} \quad S^{-1}(0, 1) = (-2, 1).$$

Mas isto nos diz de que forma S^{-1} age nos vetores da base canônica. Logo, por (15),

$$(S^{-1})_\varepsilon = \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}.$$

Invertendo esta matriz, obtemos

$$(S)_\varepsilon = \frac{1}{5} \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}.$$

Portanto,

$$S(x, y) = \frac{1}{5}(x + 2y, -2x + y),$$

que, como você pode facilmente verificar, satisfaz as condições desejadas.

3.4. Classificação dos operadores ortogonais. Neste artigo provamos o seguinte teorema.

TEOREMA 3.1. *A matriz na base canônica de um operador ortogonal do plano ou tem determinante 1 e o operador é uma rotação, ou determinante -1 e o operador é uma reflexão.*

 É importante que você seja alertado, desde já, que este resultado é falso quando se trata de operadores ortogonais em dimensões superiores a dois.

Seja T um operador do plano e seja

$$Q = (T)_\varepsilon$$

sua matriz na base canônica. Vimos, no artigo anterior, que T é ortogonal quando

$$QQ^t = I$$

em que I é a matriz identidade 2×2 . Usando as propriedades do determinante listadas no exercício 19, temos que

$$\det(QQ^t) = \det(Q) \det(Q^t) = \det(Q)^2.$$

Como a matriz identidade tem determinante igual a 1, podemos concluir que

$$\det(Q)^2 = 1;$$

onde $\det(Q) = \pm 1$. Por outro lado, se

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

então

$$QQ^t = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} a^2 + b^2 & ac + bd \\ ac + bd & c^2 + d^2 \end{bmatrix}$$

Para que esta matriz seja igual à identidade, devemos ter que

$$(20) \quad a^2 + b^2 = c^2 + d^2 = 1 \quad \text{e} \quad ac + bd = 0.$$

A partir deste ponto precisamos considerar dois casos, dependendo se a é ou não igual a zero. Se $a = 0$, então

$$bd = 0.$$

Contudo, b não pode ser nulo porque isto é incompatível com

$$a^2 + b^2 = 1,$$

uma vez que estamos supondo que $a = 0$. Logo, $d = 0$ e

$$b^2 = c^2 = 1.$$

Portanto, quando $a = 0$, as únicas possibilidades para Q são

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad \text{e} \quad \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Destas quatro matrizes, as duas primeiras descrevem reflexões, ao passo que as duas últimas descrevem rotações de ângulos $\pi/2$ e $3\pi/2$, respectivamente. Logo, o teorema é verdadeiro quando $a = 0$.

Supondo, agora, que $a \neq 0$, temos da terceira equação em (20) que

$$c = -\frac{bd}{a}.$$

Substituindo isto em $c^2 + d^2 = 1$, obtemos

$$\left(\frac{bd}{a}\right)^2 + d^2 = 1.$$

Eliminando o denominador e pondo d^2 em evidência do lado direito

$$d^2(b^2 + a^2) = a^2.$$

Levando em conta que $a^2 + b^2 = 1$, resta que $d^2 = a^2$; donde $d = \pm a$. Substituindo isto $ac = -bd$ e levando em conta que $a \neq 0$, verificamos que,

$$d = a \quad \text{e} \quad c = -b \quad \text{ou} \quad d = -a \quad \text{e} \quad c = b.$$

Logo,

$$Q = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \quad \text{ou} \quad Q = \begin{bmatrix} a & b \\ b & -a \end{bmatrix}.$$

Contudo, como $a^2 + b^2 = 1$, o vetor (a, b) tem seu vértice sobre a circunferência de centro na origem e raio um. Logo, existe um ângulo θ tal que

$$a = \cos(\theta) \quad \text{e} \quad b = \sin(\theta).$$

Levando isto em conta, as duas matrizes acima podem ser reescritas na forma

$$Q = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \quad \text{ou} \quad Q = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix}.$$

Lembrando que

$$\sin(-\theta) = -\sin(\theta) \quad \text{e que} \quad \cos(-\theta) = \cos(\theta)$$

a primeira das duas matrizes acima é igual a

$$\begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix}$$

que corresponde a uma rotação anti-horária de ângulo $-\theta$.

Para mostrar que

$$Q = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix}$$

é a matriz de uma reflexão, considere os vetores

$$u_{\pm} = Qe_1 \pm e_1 = (\cos(\theta) \pm 1, \sin(\theta)).$$

Mas,

$$Qu_{\pm} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix} \cdot \begin{bmatrix} \cos(\theta) \pm 1 \\ \sin(\theta) \end{bmatrix} = \begin{bmatrix} 1 \pm \cos(\theta) \\ \pm \sin(\theta) \end{bmatrix},$$

ao passo que,

$$\langle u_+ | u_- \rangle = (\cos(\theta) + 1)(\cos(\theta) - 1) + \sin(\theta)^2 = 0.$$

Portanto, u_+ e u_- são vetores ortogonais que satisfazem

$$Qu_+ = u_+ \quad \text{e} \quad Qu_- = -u_-,$$

o que mostra que, neste caso, Q é uma reflexão cujo espelho é a reta com vetor diretor u_+ , complementando, assim, a demonstração do teorema.

Exercícios

1. Sejam u e v vetores do plano. Use as propriedades do produto interno para calcular

$$\langle u + v | u + v \rangle, \quad \langle u - v | u - v \rangle \quad \text{e} \quad \langle u - v | u + v \rangle$$

em função de $\langle u | v \rangle$ e das normas de u e v .

2. Se $v = (2, 3)$ são as coordenadas de um vetor v em relação à base canônica, quais são as suas coordenadas relativamente à base ortonormal

$$\beta = \left\{ \frac{1}{\sqrt{5}}(1, 2), \frac{1}{\sqrt{5}}(2, -1) \right\},$$

cujos vetores também estão escritos em coordenadas na base canônica?

3. Seja ε uma base do plano formada por dois vetores unitários, e_1 e e_2 , perpendiculares entre si. Prove que todo vetor v do plano pode ser escrito na forma

$$v = \langle v | e_1 \rangle e_1 + \langle v | e_2 \rangle e_2.$$

4. Calcule o ângulo entre as retas $2x + 3y = 0$ e $5x + 2y = 0$.

5. Escreva cada o vetor $v = (2, 3)$ como combinação linear de cada um dos pares de vetores w_1 e w_2 dados abaixo:

(a) $w_1 = (1, 1)$ e $w_2 = (1, 0)$;

- (b) $w_1 = (1, 2)$ e $w_2 = (2, 1)$;
 (c) $w_1 = (7, 3)$ e $w_2 = (-14, -21)$.
6. Sejam u_1 e u_2 vetores do plano e U a matriz cuja primeira linha é u_1 e cuja segunda linha é u_2 . Prove que as seguintes afirmações são equivalentes:
 (a) u_1 e u_2 não são colineares;
 (b) $\det(U) \neq 0$.
7. Seja
- $$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$
- uma matriz 2×2 e seja $\varepsilon = \{e_1, e_2\}$ a base canônica do plano.
- (a) Calcule o vetor Ae_i .
 (b) Mostre que, se B é outra matriz 2×2 e $Av = Bv$ para todo vetor v do plano, então $A = B$.
8. Determine as matrizes que correspondem às seguintes transformações lineares:
 (a) um cisalhamento que leva a reta $x = 0$ em $y = 2x$;
 (b) uma rotação anti-horária de $\pi/6$ radianos;
 (c) uma reflexão cujo espelho é a reta $y = 2x$;
 (d) uma projeção sobre a reta $y = 3x$.
9. Em computação quântica a matriz
- $$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$
- descreve a chamada *porta de Hadamard*.
 (a) Mostre que H descreve uma reflexão do plano.
 (b) Calcule o espelho desta reflexão.
10. Seja P o operador do \mathbb{R}^2 que descreve a projeção ortogonal sobre uma reta ℓ . Determine a matriz de P na base canônica sabendo-se que $P(1, 1) = (10, 15)$.
11. Seja ρ_θ uma rotação anti-horária de um ângulo θ no plano. Para quais valores de θ a matriz de ρ_θ é simétrica?
12. Seja R uma reflexão do plano. Sabendo-se que

$$R(1, 3) = -\frac{1}{169}(241, 477),$$

determine:

- (a) a reta em torno da qual se dá esta reflexão, isto é o espelho de R ;
 (b) um vetor unitário perpendicular a esta reta;

- (c) a matriz de R .
13. Seja $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ o operador linear que consiste de uma rotação anti-horária de $\pi/3$ radianos seguida de uma reflexão que tem a reta $y = 2x$ como espelho.
- Determine a matriz de T na base canônica.
 - T é uma rotação ou uma reflexão?
14. Considere o operador linear S do plano que corresponde à reflexão R que leva o vetor $(5, 12)$ no vetor $(0, 13)$ *seguida* da rotação *horária* de $\pi/6$ que chamaremos de ρ . Sabendo-se que as coordenadas dos vetores acima foram dadas relativamente à base orthonormal $\varepsilon = \{e_1, e_2\}$ do plano formada por vetores unitários e perpendiculares entre si, determine:
- as matrizes de R e ρ relativamente à base ε ;
 - a matriz de S relativamente à base ε ;
 - a imagem do vetor $(1, 1)$ pelo operador S .
15. Determine a matriz de um operador linear $S : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ que leve o paralelogramo de vértices
- $$(0, 0), (2, 0), (2, 1), (4, 1)$$
- no quadrado de vértices
- $$(0, 0), (1, 0), (0, 1), (1, 1).$$

16. Mostre que não existe uma rotação capaz de levar o sistema de eixos definido pela base canônica $\varepsilon = \{e_1, e_2\}$ do plano no sistema de eixos definido pela base $\beta = \{-e_1, e_2\}$.
17. Dada uma matriz

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

de determinante não nulo mostre que a matriz

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

satisfaz $AA^{-1} = A^{-1}A = I$, em que I é a matriz identidade 2×2 e

$$\det(A) = a_{11}a_{22} - a_{12}a_{21}$$

é o determinante de A . A matriz A^{-1} é a *inversa* da matriz A . Quando A tem inversa, dizemos que é uma matriz *inversível*.

18. Existe alguma matriz inversível A , de tamanho 2×2 , que satisfaça à condição $A^{-1} = -A$? Dê exemplo de uma tal matriz, se a resposta for sim; caso contrário, prove que tal matriz não pode existir.
19. Sejam A e B matrizes quadradas 2×2 e λ um número real. Prove as seguintes propriedades do determinante:
 - (a) $\det(A^t) = \det(A)$;
 - (b) $\det(AB) = \det(A)\det(B)$;
 - (c) $\det(\lambda A) = \lambda^2 \det(A)$.

CAPÍTULO 2

Autovalores e autovetores

Os autovalores e autovetores de um operador linear surgem em um número enorme de aplicações, em áreas tão díspares quanto computação quântica e algoritmos de busca na web. Neste capítulo introduzimos os conceitos básicos da teoria de autovalores e autovetores no caso especial em que a matriz do operador é simétrica.

1. Autovalores e autovetores

Começaremos esta seção investigando o que acontece quando aplicamos um operador linear vezes sucessivas a um vetor. Mais precisamente, se v_0 é um vetor e T é um operador linear do plano, como se comporta a sequência de vetores $v_0, T(v_0), T^2(v_0), T^3(v_0), \dots$?

1.1. Esticando vetores. Dados um operador linear

$$T_1(x, y) = (2x, 3y),$$

e um vetor v_0 do plano, queremos determinar os vetores da sequência

$$v_0, T_1(v_0), T_1^2(v_0), T_1^3(v_0), \dots$$

Há duas escolhas de v_0 para as quais o cálculo da sequência é quase imediato. Como

$$T_1(e_1) = 2e_1 \quad \text{e} \quad T_1(e_2) = 3e_2$$

então

$$T_1^k(e_1) = 2^k e_1 \quad \text{e} \quad T_1^k(e_2) = 3^k e_2.$$

Em outras palavras, a cada iteração obtemos um vetor cada vez maior, mas com a mesma direção e sentido do vetor inicial. Vejamos, então, o que acontece quando escolhemos $v_0 = (1, 1)$. Como a primeira coordenada dobra e a segunda triplica a cada iteração, depois de k iterações teremos

$$(21) \quad T_1^k(1, 1) = (2^k, 3^k).$$

A primeira coisa a observar é que a norma do vetor cresce exponencialmente, porque

$$\|T_1^k(1, 1)\| = \sqrt{2^{2k} + 3^{2k}} > 3^k.$$

Além disso, a segunda coordenada de $T_1^k(1, 1)$ cresce mais rápido que a primeira, porque tem 3 como base da exponencial, ao passo que a base da primeira coordenada é 2. Isto

faz com que o vetor $T_1^k(1, 1)$ vá subindo na direção vertical à medida que k vai crescendo, como ilustrado na figura 1, onde foram desenhados os vetores correspondentes às primeiras cinco iterações.

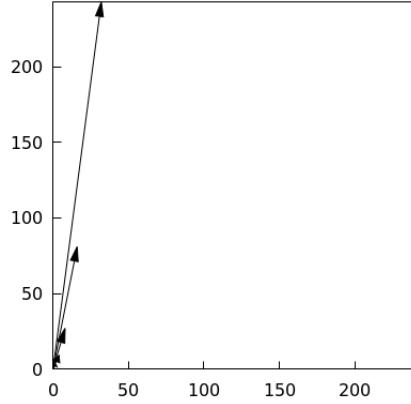


FIGURA 1. Cinco iterações de T_1 .

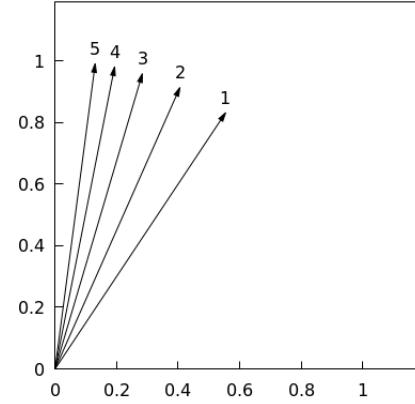


FIGURA 2. Cinco iterações normalizadas de T_1 .

Infelizmente, o aumento exponencial do comprimento do vetor torna a figura um pouco difícil de interpretar, porque os vetores resultantes das primeiras iterações ficam quase invisíveis, tal é a diferença de tamanho entre eles e $T_1^5(1, 1)$. Contudo, como o que nos interessa neste momento é a direção do vetor, substituímos, na figura 2, cada vetor por sua normalização. Como não é mais possível distinguir a que iteração um dado vetor pertence considerando apenas seu tamanho, cada vetor agora aparece rotulado pela iteração correspondente.

A figura sugere que, levando a iteração às últimas consequências, deveríamos obter o vetor $(0, 1)$. Vejamos como provar isto rigorosamente. Em primeiro lugar, temos de (21) que

$$\frac{T_1^k(1, 1)}{\|T_1^k(1, 1)\|} = \frac{1}{\sqrt{2^{2k} + 3^{2k}}} (2^k, 3^k) = \left(\sqrt{\frac{2^{2k}}{2^{2k} + 3^{2k}}}, \sqrt{\frac{3^{2k}}{2^{2k} + 3^{2k}}} \right).$$

Dividindo o numerador e o denominador de cada um dos radicais por 3^{2k} , obtemos

$$\frac{T_1^k(1, 1)}{\|T_1^k(1, 1)\|} = \left(\sqrt{\frac{(2/3)^{2k}}{(2/3)^{2k} + 1}}, \sqrt{\frac{1}{(2/3)^{2k} + 1}} \right).$$

Finalmente, como $2/3 < 1$, segue-se que

$$\lim_{k \rightarrow \infty} \frac{T_1^k(1, 1)}{\|T_1^k(1, 1)\|} = (0, 1),$$

como queríamos provar.

1.2. Tirando coelho da cartola. Para nosso próximo exemplo, utilizaremos o operador

$$(22) \quad T_2(x, y) = (117x - 144y, -144x + 33y).$$

Para evitar que a variação no comprimento do vetor torne a figura difícil de interpretar, usaremos a mesma estratégia já aplicada a T_1 , normalizando os vetores a cada iteração. Mais precisamente, os vetores na figura 3 foram calculados usando a recursão

$$u_{k+1} = T_2(u_k)/\|T_2(u_k)\|,$$

com ponto de partida $u_0 = (1, 1)/\sqrt{2}$. Na figura 3 o vetor u_k aparece rotulado por k .

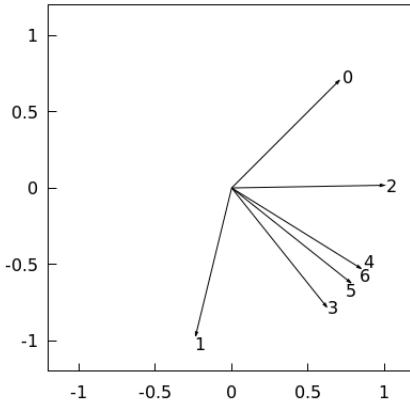


FIGURA 3. Cinco iterações normalizadas de T_2 .

A razão pela qual você não consegue distinguir u_4 de u_6 é que estão muito próximos um do outro. Para ver quão próximos, basta subtrair um do outro

$$u_4 - u_6 = (0.0429, 0.0633).$$

Aumentando o número de iterações, a diferença entre os vetores de duas iterações consecutivas vai diminuindo rapidamente. Por exemplo,

$$u_{100} - u_{99} = (3.2597 \cdot 10^{-47}, 4.3463 \cdot 10^{-47}).$$

Isto sugere que a sequência

$$u_0, u_1, u_2, \dots$$

tem um limite u_∞ bem definido quando k tende a infinito. Mas, para que isto aconteça é necessário que

$$\frac{T_2(u_\infty)}{\|T_2(u_\infty)\|} = \pm u_\infty,$$

que é equivalente a

$$(23) \quad T_2(u_\infty) = \pm \|T_2(u_\infty)\| u_\infty.$$

Tomando

$$u_{100} = (0.7999999999999869415, -0.599999999999902061)$$

como aproximação para u_∞ , teremos que $T_2(u_\infty)$ é aproximadamente igual a

$$T_2(u_{100}) = (179.9999999999970618, -134.999999999977963)$$

cuja norma é

$$224.9999999999963273.$$

Mas,

$$225 - \|T_2(u_{100})\| = 3.67 \cdot 10^{-14}$$

de modo que há uma boa chance de que o valor exato para a norma seja 225. Supondo que isto seja verdade, teremos de (23) que

$$(24) \quad T_2(u_\infty) = 225 \cdot u_\infty.$$

Supondo que $u_\infty = (x_0, y_0)$ e usando a fórmula para T_2 em (22), a equação acima nos dá

$$(117x_0 - 144y_0, -144x_0 + 33y_0) = 225(x_0, y_0).$$

Subtraindo o lado direito do lado esquerdo,

$$(-108x_0 - 144y_0, -144x_0 - 192y_0) = (0, 0),$$

que corresponde ao sistema linear

$$(25) \quad \begin{aligned} -108x_0 - 144y_0 &= 0, \\ -144x_0 - 192y_0 &= 0. \end{aligned}$$

Isolando o valor de y_0 na primeira equação

$$y_0 = -\frac{3}{4}x_0.$$

Mas substituindo isto em $-144x_0 - 192y_0$ teremos

$$-144x_0 - 192 \left(-\frac{3}{4}x_0 \right) = -144x_0 + 144x_0 = 0,$$

o que mostra que o sistema é indeterminado. Portanto, as soluções de (25) são os vetores ao longo da reta ℓ cuja equação cartesiana é

$$-108x - 144y = 0.$$

Como $(4, -3) \in \ell$, teremos que

$$u_\infty = \frac{1}{5}(4, -3),$$

pois se trata de um vetor normalizado. Encerraremos tirando um coelho da cartola: o vetor

$$u_\infty^\perp = \frac{1}{5}(3, 4),$$

é ortogonal a u_∞ e um cálculo simples mostra que

$$T_2(u_\infty^\perp) = -75u_\infty^\perp.$$

Intrigante, não? Curiosamente, o mesmo vale para T_1 , já que a sequência de vetores normalizados converge para e_2 e $T_1(e_1) = 2e_1$. Resta-nos perguntar: quão geral é o comportamento ilustrado pelos operadores T_1 e T_2 ? A resposta a esta pergunta vai nos ocupar pelo resto do capítulo.

1.3. Autovalores e autovetores. Neste artigo introduzimos a terminologia que será utilizada pelo resto do capítulo. Seja T um operador do plano. Um *autovetor* de T , associado ao *autovalor* $\lambda \in \mathbb{R}$ é um vetor não nulo v do plano que satisfaz a equação

$$T(v) = \lambda v.$$

Se μ é um número real qualquer, então

$$T(\mu v) = \mu T(v) = \mu(\lambda v) = \lambda(\mu v),$$

mostra que todo múltiplo de v também é autovetor de T associado a λ . Podemos exprimir este resultado de maneira mais geométrica dizendo que a reta ℓ cujo vetor diretor é v é *invariante* por T . Em outras palavras, ao aplicar T a um vetor de ℓ obtemos sempre vetores que pertencem a ℓ . Por causa disso sempre nos referiremos a *um*, em vez de *ao*, autovetor associado ao autovalor λ .

 Note que um autovalor pode ser nulo, mas não um autovetor; *um autovetor é, por definição, não nulo*. Um exemplo de operador que tem autovetores associados a um autovalor nulo é a projeção ortogonal sobre uma reta, porque qualquer vetor normal à reta é levado no vetor zero.

Aplicando esta terminologia aos operadores T_1 e T_2 que investigamos nos artigos 1.1 e 1.2, temos que $e_1 = (1, 0)$ e $e_2 = (0, 1)$ são autovetores do operador T_1 associados aos autovalores 3 e 2; ao passo que $(4, 3)$ e $(3, -4)$ são autovetores de T_2 associados aos autovalores 225 e -75 .

Podemos facilmente determinar os autovalores e autovetores das rotações, projeções, reflexões e cisalhamentos usando apenas sua descrição geométrica. Começando pelas rotações, é claro que não deixam nenhuma reta invariante a não ser quando o ângulo de rotação é um múltiplo inteiro de π . Mais precisamente, se o ângulo é múltiplo de 2π a rotação é o operador identidade e todos os vetores não nulos são autovetores associados ao autovalor um. Por outro lado, se o ângulo é igual a um número ímpar vezes π , então a rotação inverte o sinal de cada vetor, de modo que todos os vetores não nulos do plano são autovetores associados ao autovalor -1 . Portanto, uma rotação não tem autovetores, a não ser nestes dois casos.

No caso da projeção sobre uma reta ℓ , sabemos que nada acontece aos vetores sobre ℓ , ao passo que os vetores sobre a reta ortogonal a ℓ são todos levados no vetor nulo. Portanto, os vetores de ℓ são autovetores associados ao autovalor um e a reta ortogonal a ℓ é formada pelos autovetores associados ao zero. A análise da reflexão leva a um resultado semelhante:

os vetores do espelho são autovetores associados a 1, ao passo que os vetores ortogonais a ℓ são autovetores associados a -1 .

Deixamos o cisalhamento para o fim, porque sua análise é mais delicada. Digamos que o cisalhamento S que estamos considerando deixa fixos os múltiplos de e_1 e move e_2 para a posição $e_2 + ce_1$, para algum número real não nulo c . Claramente, os múltiplos de e_1 são autovetores associados ao autovalor 1 e a maneira como S inclina e_2 torna provável que não haja outros autovetores. Para provar isto, digamos que $v \neq 0$ seja um autovetor de S . Por definição, isto significa que existe um número real λ tal que $S(v) = \lambda v$. Escrevendo o vetor v em termos da base ortonormal formada por e_1 e e_2 , temos $v = a_1e_1 + a_2e_2$, em que $a_1, a_2 \in \mathbb{R}$. Usando a linearidade de S ,

$$S(v) = S(a_1e_1 + a_2e_2) = a_1S(e_1) + a_2S(e_2).$$

Mas estamos supondo que $S(e_1) = e_1$ e $S(e_2) = e_2 + ce_1$, donde

$$S(v) = S(a_1e_1 + a_2e_2) = a_1e_1 + a_2(e_2 + ce_1) = (a_1 + ca_2)e_1 + a_2e_2.$$

Logo, $S(v) = \lambda v$ é equivalente à igualdade

$$(a_1 + ca_2)e_1 + a_2e_2 = (\lambda a_1)e_1 + (\lambda a_2)e_2.$$

Como cada vetor tem uma única expressão em termos de uma base, concluímos que

$$a_1 + ca_2 = \lambda a_1 \quad \text{e} \quad a_2 = \lambda a_2.$$

Logo, pela segunda equação, $\lambda = 1$ ou $a_2 = 0$. Porém, substituindo $\lambda = 1$ na primeira equação, obtemos $ca_2 = 0$, donde $a_2 = 0$, pois $c \neq 0$ por hipótese. Portanto, as duas equações acima implicam que $a_2 = 0$, confirmando nossa impressão inicial de que os múltiplos não nulos de e_1 são os únicos autovetores de S .

2. Calculando autovalores e autovetores

Nesta seção veremos que há uma maneira sistemática de calcular os autovalores e autovetores de um operador a partir de sua matriz, usando determinantes e sistemas lineares.

2.1. Calculando autovalores. Seja T um operador linear do plano e seja

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

sua matriz na base canônica. A equação $T(v) = \lambda v$ usada para definir $v \neq 0$ como autovetor associado ao autovalor λ pode ser traduzida na forma

$$(26) \quad \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \cdot \begin{bmatrix} x \\ y \end{bmatrix},$$

em que representamos o vetor v pela matriz coluna $[x, y]^t$. Multiplicando as matrizes e igualando as respectivas entradas dos dois lados da equação obtemos

$$a_{11}x + a_{12}y = \lambda x \quad \text{e} \quad a_{21}x + a_{22}y = \lambda y,$$

que são equivalentes ao sistema linear

$$(27) \quad \begin{aligned} (a_{11} - \lambda)x + a_{12}y &= 0 \\ a_{21}x + (a_{22} - \lambda)y &= 0. \end{aligned}$$

Como os termos constantes são nulos em ambas as equações, o sistema tem como solução possível $x = y = 0$, que corresponde ao vetor nulo e, portanto, não nos dá nenhum autovetor de T . Logo, para que haja autovetores é necessário que o sistema (27) tenha mais de uma solução, o que só ocorre quando uma das equações é um múltiplo constante da outra. Digamos que é este o caso e que

$$(a_{11} - \lambda)x + a_{12}y = \mu(a_{21}x + (a_{22} - \lambda)y).$$

Igualando os coeficientes de x e y nos dois lados desta equação,

$$a_{11} - \lambda = \mu a_{21} \quad \text{e} \quad a_{12} = \mu(a_{22} - \lambda).$$

Segue destas equações que

$$\mu = \frac{a_{11} - \lambda}{a_{21}} = \frac{a_{12}}{a_{22} - \lambda};$$

onde

$$(a_{11} - \lambda)(a_{22} - \lambda) = a_{21}a_{12};$$

que equivale a

$$\lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{21}a_{12}) = 0.$$

Portanto, qualquer autovalor λ de T é uma raiz da equação quadrática $p_A(t) = 0$, em que

$$p_A(t) = t^2 - (a_{11} + a_{22})t + (a_{11}a_{22} - a_{21}a_{12}),$$

é o *polinômio característico* de A (e de T). Felizmente é possível usar determinantes para escrever o polinômio característico de maneira compacta na forma

$$(28) \quad p_A(t) = \det(A - tI) = \det \begin{bmatrix} a_{11} - t & a_{12} \\ a_{21} & a_{22} - t \end{bmatrix}.$$

em que I denota a matriz identidade de tamanho 2×2 .

Vejamos como usar o polinômio característico para calcular os autovalores dos operadores T_1 e T_2 definidos nos artigos 1.1 e 1.2. As matrizes destes operadores na base canônica são, respectivamente,

$$(29) \quad A_1 = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \quad \text{e} \quad A_2 = \begin{bmatrix} 117 & -144 \\ -144 & 33 \end{bmatrix}.$$

O polinômio característico de T_1 é

$$p_{A_1}(t) = \det \begin{bmatrix} 2-t & 0 \\ 0 & 3-t \end{bmatrix} = (t-2)(t-3),$$

de modo que seus autovalores são 2 e 3; já o polinômio característico de T_2 é

$$p_{A_2}(t) = \det \begin{bmatrix} 117-t & -144 \\ -144 & 33-t \end{bmatrix} = t^2 - 150t - 16875.$$

Calculando as raízes deste último polinômio, encontramos 225 e -75 , como seria de esperar do que já havíamos descoberto no artigo 1.2.

O polinômio característico também nos ajuda a entender porque, do ponto de vista algébrico, uma rotação não tem autovetores. Vimos no artigo 2.3 que a matriz de uma rotação anti-horária de ângulo θ é

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix},$$

que tem como polinômio característico

$$\det \begin{bmatrix} \cos(\theta) - t & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) - t \end{bmatrix} = t^2 - 2\cos(\theta)t + 1,$$

já que $\sin(\theta)^2 + \cos(\theta)^2 = 1$. Mas o discriminante deste polinômio do segundo grau é

$$\Delta = 4(\cos(\theta)^2 - 1)$$

que é negativo a menos que $\cos(\theta) = \pm 1$. Portanto, como nossa análise geométrica já havia mostrado, uma rotação só admite autovetores quando θ é múltiplo de π .

2.2. Calculando autovetores. Uma vez que conhecemos os autovalores de um operador, fica fácil encontrar seus autovetores. Supondo, como no artigo anterior, que A é a matriz do operador T , calculamos os autovetores de T associados ao autovalor λ simplesmente resolvendo a equação matricial

$$(A - \lambda I)X = 0,$$

em que I é a matriz identidade e $X = [x, y]^t$ é a matriz cujas entradas são as variáveis x e y .

Por exemplo, o operador

$$T_3 = (x + 2y, 2x + y),$$

cuja matriz é

$$A_3 = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

tem

$$p_3(t) = (1-t)^2 - 4$$

como polinômio característico; de modo que seus autovalores são 3 e -1 . Os autovetores associados a 3 são as soluções da equação

$$(A_3 - 3I)X = \begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

que é equivalente ao sistema linear

$$\begin{aligned} -2x + 2y &= 0; \\ 2x - 2y &= 0. \end{aligned}$$

Como seria de esperar, pela análise no artigo anterior, a segunda equação é um múltiplo da primeira. Portanto, este sistema tem infinitas soluções, dadas $y = x$, e que correspondem aos vetores

$$(x, y) = (x, x) = x(1, 1).$$

Logo, qualquer múltiplo não nulo de $(1, 1)$ é autovetor de T_3 associado ao autovalor 3. Já os autovetores associados a -1 são as soluções de

$$(A_3 + I)X = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

que equivale ao sistema linear

$$\begin{aligned} 2x + 2y &= 0 \\ 2x + 2y &= 0. \end{aligned}$$

Desta vez as duas equações do sistema são iguais e delas obtemos $y = -x$. Logo as soluções do segundo sistema linear são os vetores

$$(x, y) = (x, -x) = x(1, -1).$$

Assim, os autovetores de T_3 associados ao autovalor -1 são os múltiplos não nulos de $(1, -1)$. Note que os dois autovetores que obtivemos são ortogonais, exatamente como já havia acontecido com os autovetores dos operadores T_1 e T_2 na seção 1.

A esta altura você talvez tenha concluído que os autovetores de um operador do plano são sempre ortogonais. Infelizmente isto é falso, como mostra o próximo exemplo. Seja T_4 o operador do plano definido por

$$T_4(x, y) = (10x + 6y, -12x - 7y),$$

cuja matriz na base canônica é

$$A_4 = \begin{bmatrix} 10 & 6 \\ -12 & -7 \end{bmatrix}.$$

Este operador tem polinômio característico igual a

$$\det \begin{bmatrix} 10 - t & 6 \\ -12 & -7 - t \end{bmatrix} = t^2 - 3t + 2.$$

Resolvendo a equação quadrática $t^2 - 3t + 2 = 0$, verificamos que os autovalores de T_4 são iguais a 2 e 1. Para achar os autovetores de T_4 associados a 2 precisamos encontrar as soluções da equação matricial

$$(A_4 - 2I) = \begin{bmatrix} 8 & 6 \\ -12 & -9 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

que é equivalente ao sistema linear

$$\begin{aligned} 8x + 6y &= 0, \\ -12x - 9y &= 0. \end{aligned}$$

Como estas duas equações são múltiplas uma da outra, obtemos $y = -4x/3$ a partir de qualquer uma delas. Logo, as soluções do sistema são dadas pelos vetores

$$(x, y) = \left(x, \frac{-4}{3}x \right) = \frac{1}{3}x(3, -4).$$

Assim, os autovetores de T_4 associados ao autovalor 2 são os múltiplos não nulos de $(3, -4)$. Procedendo de maneira análoga, os autovetores de 1 são as soluções de

$$(A_4 - I) = \begin{bmatrix} 9 & 6 \\ -12 & -8 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

que equivale ao sistema linear

$$\begin{aligned} 9x + 6y &= 0 \\ -12x - 8y &= 0. \end{aligned}$$

Como as equações são múltiplas uma da outra, os vetores (x, y) que desejamos têm que satisfazer $y = -3x/2$; donde

$$(x, y) = \frac{x}{2}(2, -3);$$

o que mostra que os autovetores de T_4 , associados ao autovalor 1, são os múltiplos não nulos de $(2, -3)$. Finalmente, calculando o produto interno dos autovetores associados a 2 e 1 que encontramos teremos

$$\langle (3, -4) | (2, -3) \rangle = 3 \cdot 2 + (-4) \cdot (-3) = 18$$

de forma que estes autovetores não são ortogonais; ao contrário do que havia acontecido com todos os operadores que havíamos usado como exemplo até aqui.

Evidentemente, se formos capazes de determinar *a priori* que os autovetores de um dado operador do plano são ortogonais, podemos calcular um deles e com isso achar o

segundo sem nenhum cálculo adicional. Na próxima seção veremos que os operadores com esta propriedade são fáceis de caracterizar. Infelizmente, estes operadores constituem uma minoria ínfima do conjunto de todos os operadores lineares do plano. Contudo, apesar de serem minoria, estes operadores desempenham um papel extremamente importante, aparecendo em inúmeras aplicações da álgebra linear a outras áreas da matemática, à física e à engenharia.

3. O teorema espectral

Ao final do artigo 1.2 tiramos subitamente um coelho da cartola, afirmando que um vetor ortogonal ao autovetor de T_2 associado ao autovalor 225, que já havíamos encontrado, também era autovetor, só que associado ao autovalor -75 . Na seção anterior, ao mesmo tempo que vimos um terceiro exemplo de operador com a mesma propriedade, descobrimos que ela se aplica a muito poucos operadores, ainda que estes tenham um papel importante em muitas aplicações. Nesta seção caracterizamos os operadores do plano cujos autovetores, correspondentes a autovalores distintos, são ortogonais. Ao final da seção, formalizaremos o método iterativo utilizado para encontrar autovetores nos artigos 1.1 e 1.2 da seção anterior.

3.1. Teorema espectral. Já sabemos que há operadores que não têm nenhum autovetor, como é o caso de todas as rotações cujo ângulo não é múltiplo de π . Contudo, há famílias inteiras de operadores que sempre têm autovetores, a mais importante delas, por causa de suas inúmeras aplicações, é a dos operadores autoadjuntos, que têm a propriedade adicional de que os autovetores associados a autovalores diferentes são sempre ortogonais.

Dizemos que um operador linear T do plano é *autoadjunto* se

$$(30) \quad \langle T(u) | v \rangle = \langle u | T(v) \rangle,$$

quaisquer que sejam os vetores u e v do plano. O problema desta definição é que, ao menos em princípio, precisamos testá-la para todas as escolhas possíveis de vetores u e v do plano, o que é evidentemente impossível. Felizmente, há uma tradução matricial da condição (30) que é extremamente fácil de verificar. Digamos que A é a matriz de um operador autoadjunto T e que u e v são vetores do plano que identificaremos com suas representações matriciais em termos de matrizes coluna. Usando (17), podemos reescrever a condição (30) na forma

$$(Au)^t v = u^t (Av)$$

quaisquer que sejam as matrizes u e v de tamanho 2×1 . Lembrando que $(Au)^t = u^t A^t$, a equação acima toma a forma

$$u^t A^t v = u^t Av.$$

Mas, se isto vale para todas as escolhas de u e v , então $A^t = A$; veja exercício 7 da página 33. Escrevendo

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \text{temos que} \quad A^T = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix},$$

de modo que $A^t = A$ equivale a dizer que $a_{21} = a_{12}$; isto é, que

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix}.$$

Estas matrizes são chamadas de *simétricas*, porque suas entradas dos dois lados da diagonal principal são iguais. Portanto, o cálculo acima mostra que um operador é *autoadjunto* se, e somente se, sua matriz, relativamente à base canônica, é simétrica.

Dentre as classes de operadores que já estudamos, as projeções e as reflexões são operadores autoadjuntos. Por exemplo, se u é um vetor unitário então, como vimos no artigo 2.2 do capítulo 1, a matriz da reflexão cujo espelho é a reta de vetor normal unitário u é igual a

$$R = I - 2uu^t.$$

Mas,

$$(uu^t)^t = (u^t)^tu^t = uu^t,$$

onde

$$R^t = (I - 2uu^t)^t = I^t - 2(uu^t)^t = I - 2uu^t = R,$$

pois a transposta de uma soma de matrizes é a soma das transpostas destas matrizes. Portanto, a matriz de uma reflexão relativamente à base canônica é simétrica, como desejávamos mostrar.

TEOREMA ESPECTRAL (em dimensão dois). *Um operador autoadjunto que não é um múltiplo constante da identidade tem dois autovalores reais distintos aos quais estão associados autovetores ortogonais.*

DEMONSTRAÇÃO. Suponhamos que T seja um operador autoadjunto. Há duas coisas a provar. A primeira é que T tem dois autovalores reais distintos; a segunda que os autovetores aos quais estes autovalores estão associados são ortogonais. Começaremos analisando os autovalores.

AUTOVALORES DE T : Suponhamos que a matriz de T relativamente à base canônica seja

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix}.$$

Para calcular os autovalores de T temos que achar as raízes do polinômio

$$p_A(t) = \det(A - tI) = \det \begin{bmatrix} a_{11} - t & a_{12} \\ a_{12} & a_{22} - t \end{bmatrix} = t^2 - (a_{11} + a_{22})t + (a_{11}a_{22} - a_{12}^2),$$

cujo discriminante é igual a

$$\Delta = (a_{11} + a_{22})^2 - 4(a_{11}a_{22} - a_{12}^2) = (a_{11} - a_{22})^2 + 4a_{12}^2.$$

Logo, Δ é um número real positivo, já que é uma soma de quadrados. Portanto, as duas raízes do polinômio característico de A têm que ser ambas reais. Além disso, para que as raízes sejam iguais é necessário que

$$\Delta = (a_{11} - a_{22})^2 + 4a_{12}^2 = 0,$$

o que só é possível se $a_{11} = a_{22}$ e $a_{12} = 0$. Mas, substituindo estas condições em A , obtemos

$$A = \begin{bmatrix} a_{11} & 0 \\ 0 & a_{11} \end{bmatrix} = a_{11}I.$$

Logo, os autovalores de T só podem ser iguais quando $T = a_{11} \text{id}$, em que id denota o operador identidade.

AUTOVETORES DE T : Sejam $\lambda_1 \neq \lambda_2$ os autovalores de T e sejam u_1 e u_2 autovetores associados a, respectivamente, λ_1 e λ_2 . Por (30),

$$\langle T(u_1) | u_2 \rangle = \langle u_1 | T(u_2) \rangle.$$

Levando em conta que u_1 e u_2 são autovetores,

$$\langle \lambda_1 u_1 | u_2 \rangle = \langle u_1 | \lambda_2 u_2 \rangle,$$

que equivale a

$$\lambda_1 \langle u_1 | u_2 \rangle = \lambda_2 \langle u_1 | u_2 \rangle.$$

Subtraindo o lado direito da equação do seu lado esquerdo,

$$(\lambda_1 - \lambda_2) \langle u_1 | u_2 \rangle = 0.$$

Como estamos supondo que $\lambda_1 \neq \lambda_2$, esta última equação implica que $\langle u_1 | u_2 \rangle = 0$, mostrando que u_1 e u_2 são realmente ortogonais. \square

Uma das consequências do teorema espectral é que sempre podemos escolher uma base do plano na qual um operador autoadjunto T é diagonal. Para isso, basta calcular os autovetores de T , normalizá-los e tomar o conjunto formado pelos dois autovetores como base ortonormal do plano. Infelizmente, há várias situações em que esta estratégia não é viável. Basta, por exemplo, que o problema que estamos resolvendo envolva mais de um operador autoadjunto, porque, em geral, não é possível escolher uma base que diagonalize mais de um operador ao mesmo tempo; veja exercícios 12 e 13. A saída para o dilema é descobrir como transformar as coordenadas de um vetor relativamente à base canônica nas

coordenadas do mesmo vetor em outra base ortonormal qualquer. Veremos como fazer isto nos próximos artigos.

3.2. Mudança de base de vetores. Sejam $\varepsilon = \{e_1, e_2\}$ a base canônica e $\beta = \{u_1, u_2\}$ uma outra base ortonormal cujos vetores são

$$(31) \quad u_1 = (a_{11}, a_{21}) \quad \text{e} \quad u_2 = (a_{12}, a_{22}).$$

Se as coordenadas de um vetor v relativamente à base β forem

$$(v)_\beta = (c_1, c_2)$$

então, pela definição de coordenadas

$$v = c_1 u_1 + c_2 u_2.$$

Substituindo as coordenadas de u_1 e u_2 dadas em (31) na expressão acima, obtemos as coordenadas de v relativamente à base canônica,

$$v = c_1(a_{11}, a_{21}) + c_2(a_{12}, a_{22}) = (a_{11}c_1 + c_2a_{12}, a_{21}c_1 + a_{22}c_2),$$

A matriz coluna correspondente a este último vetor é

$$(32) \quad \underbrace{\begin{bmatrix} a_{11}c_1 + c_2a_{12} \\ a_{21}c_1 + c_2a_{22} \end{bmatrix}}_{(v)_\varepsilon} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \underbrace{\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}}_{(v)_\beta}.$$

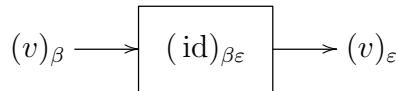
A matriz quadrada

$$(\text{id})_{\beta\varepsilon} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

é conhecida como a *matriz de mudança de base*, porque ela converte as coordenadas de um vetor na base β em suas coordenadas na base canônica. Usando esta notação, podemos reescrever (32) na forma

$$(33) \quad (v)_\varepsilon = (\text{id})_{\beta\varepsilon} \cdot (v)_\beta.$$

Usaremos o diagrama abaixo como ilustração desta equação.



Antes mesmo de darmos um exemplo é importante fazer alguns comentários sobre a matriz $(\text{id})_{\beta\varepsilon}$. O primeiro é que a razão para a estranha notação será eventualmente dada no artigo 3.2 do capítulo 7. O segundo é que a ordem das bases que servem de índice a (id) deve obedecer à ordem em que as coordenadas estão sendo transformadas: em nosso caso coordenadas na base β são convertidas em coordenadas na base canônica. Finalmente, nas

colunas de $(\text{id})_{\beta\varepsilon}$ aparecem os vetores coluna correspondentes às coordenadas de u_1 e u_2 na base canônica; isto é,

$$(34) \quad (\text{id})_{\beta\varepsilon} = \begin{bmatrix} | & | \\ (u_1)_\varepsilon & (u_2)_\varepsilon \\ | & | \end{bmatrix}.$$

Assim, quando

$$(35) \quad \beta = \left\{ \frac{1}{\sqrt{2}}(1, 1), \frac{1}{\sqrt{2}}(-1, 1) \right\}$$

a matriz de mudança de base que converte coordenadas na base β em coordenadas na base canônica é

$$(\text{id})_{\beta\varepsilon} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

Assim, se as coordenadas de v na base β são $(v)_\beta = (1, 2)$, então suas coordenadas na base canônica serão

$$(v)_\varepsilon = (\text{id})_{\beta\varepsilon}(v)_\beta = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 3 \end{bmatrix}.$$

Voltando à análise do problema geral, digamos que $\beta = \{u_1, u_2\}$ seja uma base ortonormal qualquer do plano. Usando u_1 e u_2 para representar as matrizes coluna que contêm as coordenadas destes vetores na base canônica, a fórmula (34) nos dá

$$(36) \quad (\text{id})_{\beta\varepsilon}^t (\text{id})_{\beta\varepsilon} = \begin{bmatrix} - & u_1^t & - \\ - & u_2^t & - \end{bmatrix} \begin{bmatrix} | & | \\ u_1 & u_2 \\ | & | \end{bmatrix} = \begin{bmatrix} u_1^t u_1 & u_1^t u_2 \\ u_2^t u_1 & u_2^t u_2 \end{bmatrix}.$$

Contudo, como β é uma base ortonormal, os vetores u_1 e u_2 satisfazem

$$u_1^t u_2 = \langle u_1 | u_2 \rangle = 0 \quad \text{e} \quad u_i^t u_i = \langle u_i | u_i \rangle = \|u_i\|^2 = 1,$$

para $i = 1, 2$. Substituindo isto em (36), obtemos

$$(37) \quad (\text{id})_{\beta\varepsilon}^t (\text{id})_{\beta\varepsilon} = I,$$

em que I representa a matriz identidade 2×2 . Usando a terminologia do artigo 3.3 do capítulo 1, provamos que *a matriz de mudança de base entre duas bases ortonormais é sempre ortogonal*.

O que fizemos até aqui nos mostra como usar matrizes para transformar as coordenadas de um vetor na base β em suas coordenadas na base canônica: mas, e se precisássemos

fazer o contrário? Mais precisamente, qual a matriz que transforma coordenadas de vetores na base canônica nas coordenadas do mesmo vetor na base β ? Para responder a esta pergunta, lembre-se que, pela fórmula (33),

$$(v)_\varepsilon = (\text{id})_{\beta\varepsilon} \cdot (v)_\beta.$$

Mas, multiplicando esta equação à esquerda pela transposta de $(\text{id})_{\beta\varepsilon}$,

$$(\text{id})_{\beta\varepsilon}^t (v)_\varepsilon = (\text{id})_{\beta\varepsilon}^t (\text{id})_{\beta\varepsilon} \cdot (v)_\beta.$$

Lembrando que $(\text{id})_{\beta\varepsilon}$ é ortogonal, podemos simplificar o lado direito, obtendo

$$(\text{id})_{\beta\varepsilon}^t (v)_\varepsilon = (v)_\beta;$$

o que mostra que a matriz $(\text{id})_{\varepsilon\beta}$ que efetua a mudança de coordenadas na base canônica para coordenadas na base β satisfaz

$$(\text{id})_{\varepsilon\beta} = (\text{id})_{\beta\varepsilon}^t.$$

Assim quando β é a base descrita em (35),

$$(\text{id})_{\varepsilon\beta} = (\text{id})_{\beta\varepsilon}^t = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

Podemos resumir o que vimos neste artigo no seguinte teorema.

TEOREMA 3.1. *Sejam ε a base canônica e β uma outra base ortonormal do plano. A matriz de mudança de base $(\text{id})_{\beta\varepsilon}$, obtida escrevendo em suas colunas as coordenadas dos vetores de β relativamente à base canônica satisfaz:*

- (1) $(v)_\varepsilon = (\text{id})_{\beta\varepsilon}(v)_\beta$;
- (2) $(\text{id})_{\varepsilon\beta} = (\text{id})_{\beta\varepsilon}^t$;
- (3) $(\text{id})_{\beta\varepsilon}^t (\text{id})_{\beta\varepsilon} = I$.

3.3. Mudança de base de operadores. No artigo anterior, vimos como transformar as coordenadas de um vetor entre a base canônica e uma outra base ortonormal; nesta seção veremos como estender isto às matrizes dos operadores lineares.

Antes de começar, lembre-se que não há nada de especial sobre a base canônica, se comparada a outras bases ortonormais do plano, exceto o fato de termos decidido utilizá-la como a base padrão, em relação à qual escreveríamos as coordenadas de todos os vetores do plano, *salvo menção explícita do contrário*. Portanto, podemos aplicar a qualquer base ortonormal do plano tudo o que aprendemos sobre a matriz de um operador linear no artigo 3.1 do capítulo 1.

Digamos, então, que $\beta = \{u_1, u_2\}$ é uma base ortonormal e que T é um operador do plano. Supondo que conhecemos $(Tu_1)_\beta$ e $(Tu_2)_\beta$, temos pela fórmula (14) da página 22

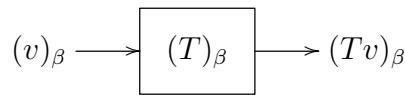
que

$$(38) \quad (T)_\beta = \begin{bmatrix} & & \\ | & & | \\ (Tu_1)_\beta & (Tu_2)_\beta \\ | & & | \end{bmatrix}.$$

Assim,

$$(39) \quad (Tv)_\beta = (T)_\beta(v)_\beta$$

que ilustramos no seguinte diagrama.



Você deve ter notado que há uma inversão da direção em que os diagramas e as fórmulas são lidos: enquanto o diagrama é lido da esquerda para a direita, a fórmula é lida da direita para a esquerda. Isto ocorre porque, por uma convenção quase que universalmente adotada, o argumento de uma função se escreve à direita do símbolo que designa o operador; em outras palavras, estamos acostumados a escrever $T(u)$ e não $(u)T$. Já nos deparamos com o mesmo fenômeno na página 26.

Vejamos como aplicar isto a um operador autoadjunto T que tem u_1 e u_2 como auto-vetores unitários associados aos autovalores λ_1 e λ_2 . Por definição,

$$T(u_1) = \lambda_1 u_1 \quad \text{e} \quad T(u_2) = \lambda_2 u_2.$$

Reescrevendo em termos de coordenadas na base β ,

$$(T(u_1))_\beta = (\lambda_1, 0) \quad \text{e} \quad (T(u_2))_\beta = (0, \lambda_2),$$

pois

$$\lambda_1 \cdot u_1 = \lambda_1 \cdot u_1 + 0 \cdot u_2 \quad \text{e} \quad \lambda_2 \cdot u_2 = 0 \cdot u_1 + \lambda_2 \cdot u_2.$$

Assim, aplicando à base β e ao operador T a fórmula (38), obtemos

$$(T)_\beta = \begin{bmatrix} & & \\ | & & | \\ (T(u_1))_\beta & (T(u_2))_\beta \\ | & & | \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix},$$

que é uma matriz diagonal.

O próximo passo consiste em descobrir como proceder para calcular $(T)_\varepsilon$, supondo que $(T)_\beta$ é conhecida. A primeira coisa a observar é que isto só é possível se conhecermos as coordenadas dos vetores de β relativamente a ε , ou vice-versa. Analisaremos o primeiro caso porque é o mais comum. Como no artigo 3.2, escreveremos u_1 e u_2 em vez de $(u_1)_\varepsilon$

e $(u_2)_\varepsilon$ para facilitar a notação, mas não esqueça que, ao fazer isto, estamos considerando u_1 e u_2 como matrizes coluna. Sob esta hipótese, as matrizes

$$(\text{id})_{\beta\varepsilon} = \begin{bmatrix} | & | \\ u_1 & u_2 \\ | & | \end{bmatrix} \quad \text{e} \quad (\text{id})_{\varepsilon\beta} = (\text{id})_{\beta\varepsilon}^t$$

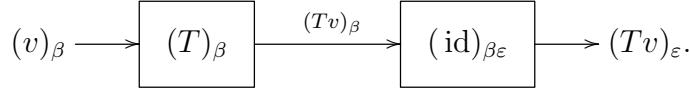
são conhecidas. Para encontrar a fórmula desejada para $(T)_\varepsilon$, utilizaremos os diagramas como se fossem blocos de código em uma linguagem de programação visual. Contudo, como em toda linguagem de programação, há uma restrição sobre o uso dos blocos que precisa ser conhecida:

só podemos ligar um bloco a outro se as coordenadas do vetor na saída do primeiro bloco estão escritas na mesma base que o vetor na entrada do segundo bloco.

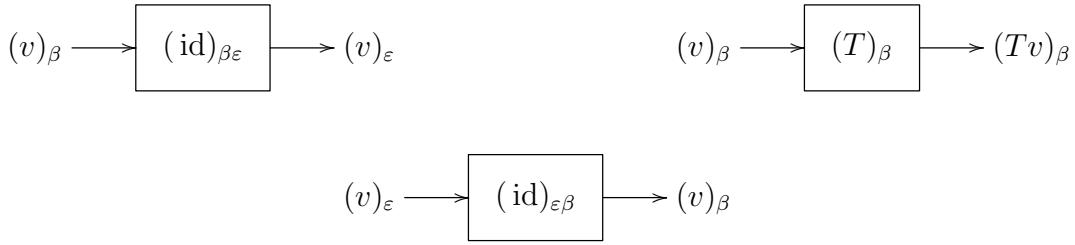
Denotando por v um vetor do plano, isto significa que os diagramas



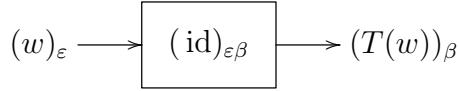
não podem ser ligados na ordem em que aparecem, mas sim na ordem oposta, como indicado abaixo.



Para podermos relacionar $(T)_\beta$ a $(T)_\varepsilon$, precisamos encaixar os diagramas



de modo que, quando aplicados a um vetor w , tenham o mesmo efeito que



Contudo, o único dos três blocos que aplica T recebe e retorna coordenadas na base β . Por isso, antes de usá-lo precisamos converter as coordenadas de w de ε para β e, depois de usá-lo, precisamos aplicar o processo inverso às coordenadas de Tw na base β . Isto sugere

que a ligação dos blocos deve ser feita como no seguinte diagrama.

$$(w)_\varepsilon \longrightarrow \boxed{(\text{id})_{\varepsilon\beta}} \xrightarrow{(w)_\beta} \boxed{(T)_\beta} \xrightarrow{(Tw)_\beta} \boxed{(\text{id})_{\beta\varepsilon}} \longrightarrow (Tw)_\varepsilon$$

Lembrando que a fórmula correspondente a um diagrama é lida da direita para a esquerda, obtemos

$$(40) \quad (\text{id})_{\beta\varepsilon}(T)_\beta(\text{id})_{\varepsilon\beta}(w)_\varepsilon = (T)_\varepsilon(w)_\varepsilon.$$

Porém, pelo exercício 7 da página 33, isto significa que

$$(41) \quad (\text{id})_{\beta\varepsilon}(T)_\beta(\text{id})_{\varepsilon\beta} = (T)_\varepsilon.$$

Um exemplo deve ajudar a entender melhor como esta fórmula é utilizada. Já sabemos que, pelo teorema espectral, todo operador autoadjunto admite uma base ortonormal de autovetores. Suponhamos, então, que T é um operador autoadjunto e que $(1, 3)$ e $(-3, 1)$ são autovetores de T associados, respectivamente, aos autovalores 2 e 6. Normalizando os autovetores de T , obtemos

$$u_1 = \frac{(1, 3)}{\sqrt{10}} \quad \text{e} \quad u_2 = \frac{(-3, 1)}{\sqrt{10}},$$

de modo que $\beta = \{u_1, u_2\}$ é uma base ortonormal do plano. Além disso, como estamos supondo que

$$T(u_1) = 2u_1 \quad \text{e} \quad T(u_2) = 6u_2$$

temos que

$$(T)_\beta = \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix}.$$

Por outro lado, como as colunas de $(\text{id})_{\beta\varepsilon}$ correspondem às matrizes que representam os vetores u_1 e u_2 , temos

$$(\text{id})_{\beta\varepsilon} = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix};$$

onde,

$$(\text{id})_{\varepsilon\beta} = (\text{id})_{\beta\varepsilon}^t = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & 3 \\ -3 & 1 \end{bmatrix}.$$

Finalmente, para encontrar a matriz de T na base canônica basta multiplicar estas matrizes na ordem estabelecida na fórmula (40):

$$\underbrace{\frac{1}{\sqrt{10}} \begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix}}_{(\text{id})_{\beta\varepsilon}} \cdot \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix}}_{(T)_\beta} \cdot \underbrace{\frac{1}{\sqrt{10}} \begin{bmatrix} 1 & 3 \\ -3 & 1 \end{bmatrix}}_{(\text{id})_{\varepsilon\beta}} = \underbrace{\frac{1}{5} \begin{bmatrix} 28 & -6 \\ -6 & 12 \end{bmatrix}}_{(T)_\varepsilon}.$$

Nossa linguagem de diagramas é muito sugestiva e ajuda a lembrar das fórmulas, mas não chega a ser uma demonstração de (40); por isso encerraremos o artigo, provando que a fórmula (40) está correta. Para isso, basta aplicarmos a matriz

$$(\text{id})_{\beta\varepsilon}(T)_\beta(\text{id})_{\varepsilon\beta}$$

às coordenadas de um vetor w na base canônica e ver o que acontece. Mas, por (33),

$$(w)_\varepsilon = (\text{id})_{\beta\varepsilon}(w)_\beta;$$

onde

$$(\text{id})_{\beta\varepsilon}(T)_\beta(\text{id})_{\varepsilon\beta}(w)_\varepsilon = (\text{id})_{\beta\varepsilon}(T)_\beta(w)_\beta.$$

Contudo, $(T)_\beta(v)_\beta = (Tv)_\beta$, o que nos dá

$$(\text{id})_{\beta\varepsilon}(T)_\beta(\text{id})_{\varepsilon\beta}(w)_\varepsilon = (\text{id})_{\beta\varepsilon}(Tw)_\beta.$$

Finalmente, $(\text{id})_{\beta\varepsilon}(v)_\beta = (v)_\varepsilon$, qualquer que seja o vetor $v \in \mathbb{R}^2$; de modo que

$$(\text{id})_{\beta\varepsilon}(T)_\beta(\text{id})_{\varepsilon\beta}(w)_\varepsilon = (\text{id})_{\beta\varepsilon}(Tw)_\beta = (Tw)_\varepsilon,$$

como desejávamos provar.

3.4. Exemplos. Neste capítulo ilustramos algumas aplicações geométricas simples de autovalores e autovetores.

PROBLEMA 3.2. *Determine um retângulo que seja levado em um quadrado de lado 1 pelo operador*

$$T(x, y) = \frac{1}{13}(83x - 12y, -12x + 73y).$$

Antes de começar a fazer contas, convém expor a estratégia a ser utilizada. Como esta matriz é simétrica, o operador T é autoadjunto. Logo, pelo teorema espectral, T tem autovetores ortogonais u_1 e u_2 , associados a autovalores λ_1 e λ_2 . Assim,

$$T(u_1) = \lambda_1 u_1 \quad \text{e} \quad T(u_2) = \lambda_2 u_2.$$

Sem perda de generalidade, podemos supor que u_1 e u_2 são vetores unitários. Com isso, as extremidades dos vetores

$$0, \quad u_1, \quad u_2 \quad \text{e} \quad u_1 + u_2$$

são vértices de um quadrado Q de lado 1. A ideia é, então, encontrar vetores que são levados por T em u_1 e u_2 . Como o efeito de T sobre estes vetores consiste simplesmente em uma mudança de escala, podemos partir de vetores que sejam múltiplos de u_1 e u_2 . De fato,

$$T\left(\frac{u_1}{\lambda_1}\right) = \lambda_1\left(\frac{u_1}{\lambda_1}\right) = u_1 \quad \text{e} \quad T\left(\frac{u_2}{\lambda_2}\right) = \lambda_2\left(\frac{u_2}{\lambda_2}\right) = u_2.$$

Portanto, os vértices dos vetores

$$0, \quad \frac{u_1}{\lambda_1}, \quad \frac{u_2}{\lambda_2} \quad \text{e} \quad \frac{u_1}{\lambda_1} + \frac{u_2}{\lambda_2}$$

formam um retângulo que é levado por T em um quadrado de lado 1.

Para poder implementar esta estratégia, precisamos calcular os autovalores e autovetores de T mas, para isto, precisamos da matriz

$$A = (T)_\varepsilon = \frac{1}{13} \begin{bmatrix} 83 & -12 \\ -12 & 73 \end{bmatrix}$$

que representa o operador T na base canônica. O polinômio característico de A é

$$p_A(t) = \det(A - tI) = \det \frac{1}{13} \begin{bmatrix} 83 - 13t & -12 \\ -12 & 73 - 13t \end{bmatrix} = t^2 - 12t + 35,$$

cujas raízes são 5 e 7. Obtivemos, assim, os autovalores de T . Para achar os autovetores precisamos resolver os sistemas lineares resultantes das equações matriciais

$$(A - 5I)X = 0 \quad \text{e} \quad (A - 7I)X = 0.$$

As equações no primeiro sistema são

$$\frac{18x}{13} - \frac{12y}{13} = 0 \quad \text{e} \quad -\frac{12x}{13} + \frac{8y}{13} = 0.$$

Como, multiplicando a primeira equação por $-2/3$ obtemos a segunda, basta considerar a primeira delas, que nos permite escrever $y = 3x/2$. Logo, os autovetores de T associados a 5 são da forma

$$(x, y) = \frac{x}{2} (2, 3).$$

Normalizando, obtemos

$$u_1 = \frac{1}{\sqrt{13}} (2, 3).$$

Como, pelo teorema espectral, um autovetor de 7 tem que ser ortogonal a um autovetor de 5, podemos afirmar que o vetor unitário

$$u_2 = \frac{1}{\sqrt{13}} (3, -2)$$

é autovetor de 7. Com isto o quadrado desejado terá por vértices as extremidades dos vetores

$$(0, 0), \quad \frac{1}{\sqrt{13}} (2, 3), \quad \frac{1}{\sqrt{13}} (3, -2) \quad \text{e} \quad \frac{1}{\sqrt{13}} (5, 1),$$

ao passo que os vértices do retângulo serão

$$(0, 0), \quad \frac{1}{5\sqrt{13}} (2, 3), \quad \frac{1}{7\sqrt{13}} (3, -2) \quad \text{e} \quad \frac{1}{35\sqrt{13}} (29, 11).$$

PROBLEMA 3.3. Determine um triângulo retângulo que seja levado em um triângulo isósceles de área 5200 pelo operador T cuja matriz na base canônica é

$$A = \begin{bmatrix} 4 & 6 \\ 6 & -1 \end{bmatrix}$$

A ideia é construir o triângulo a partir dos autovetores da matriz, cujo polinômio característico é

$$\det \begin{bmatrix} 4-t & 6 \\ 6 & -1-t \end{bmatrix} = t^2 - 3t - 40.$$

Resolvendo a equação $t^2 - 3t - 40 = 0$, descobrimos que os autovalores de T são -5 e 8 . Para achar os autovetores associados a -5 precisamos resolver a equação matricial

$$\begin{bmatrix} 9 & 6 \\ 6 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Como as duas equações do sistema são equivalentes, basta expressar y em função de x em uma delas,

$$y = -\frac{3}{2}x.$$

Logo, um autovetor associado ao autovalor -5 é $v = (2, -3)$. Portanto, pelo teorema espectral, o vetor $v^\perp = (3, 2)$, que é ortogonal a $(2, -3)$, será um autovetor associado ao autovalor 8 . Com isso, os catetos do triângulo retângulo que desejamos construir corresponderão aos vetores

$$aT(v) = -5av \quad \text{e} \quad bT(v^\perp) = 8bv^\perp$$

em que a e b são números reais positivos a serem escolhidos. Como os catetos são perpendiculares entre si, a área do triângulo formado por esses dois vetores será

$$\frac{1}{2} \| -5av \| \| 8bv^\perp \| = 20ab \| v \| \| v^\perp \| = 20ab \| v \|^2 = 260ab.$$

Igualando isto a 5200, obtemos $ab = 20$. Como o problema também requer que o triângulo seja isósceles, é necessário que

$$\| -5av \| = \| 8bv^\perp \|;$$

isto é, que $5a = 8b$, porque os dois vetores têm a mesma norma. Substituindo isto em $ab = 20$, obtemos

$$\frac{5}{8}a^2 = 20$$

onde $a = 4\sqrt{2}$, porque escolhemos tomar a e b como números positivos. Assim, os lados do triângulo retângulo levado em um triângulo isósceles de área 5200 serão

$$w = 4\sqrt{2}(2, -3) \quad \text{e} \quad w^\perp = \frac{5\sqrt{2}}{2}(3, 2),$$

3.5. Método da potência. Encerraremos a seção analisando em detalhes o processo iterativo usado nos artigos 1.1 e 1.2 para achar aproximações de autovetores. Embora este procedimento funcione para quaisquer operadores, vamos restringir nossa análise aos operadores autoadjuntos.

Começamos recordando o procedimento, na forma em que foi usado no artigo 1.2. Seja T um operador autoadjunto do plano, cujos autovalores não são nulos. Dado um vetor unitário u_0 , definimos uma sequência de vetores u_0, u_1, \dots recursivamente por

$$(42) \quad u_i = \frac{T(u_{i-1})}{\|T(u_{i-1})\|}.$$

Por exemplo, T_3 foi definido como o operador que tem

$$A_3 = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

por matriz na base canônica. Começando a iterar a partir de $u_0 = (1, 0)$, as três primeiras iterações produzirão os vetores

$$u_1 = \frac{1}{\sqrt{5}}(1, 2), \quad u_2 = \frac{1}{\sqrt{41}}(5, 4) \quad \text{e} \quad u_3 = \frac{1}{365}(13, 14).$$

Continuando, assim, verificamos que o décimo vetor da sequência será

$$u_{10} = \frac{1}{1743392201}(29525, 29524),$$

que está próximo de ter as duas entradas iguais. Para ver quanto próximo, basta dividir a primeira coordenada pela segunda e expressar o resultado em ponto flutuante,

$$\frac{29525}{29524} = 1.0000338707.$$

Chegamos, assim, por outro caminho, à conclusão que já havíamos obtido de maneira algébrica no artigo 2.2; isto é, que os múltiplos de $(1, 1)$ são os autovetores associados ao autovalor 3, que é o maior dos dois autovalores de T .

O algoritmo definido pela iteração (42) é conhecido como *método da potência*, já que é baseado no cálculo de potências sucessivas do operador. Nossa próximo teorema afirma que, sob certas condições, este algoritmo sempre converge para um autovetor. Contudo, antes de provar o teorema, precisamos de um resultado auxiliar que será bastante útil ao longo de todo o livro.

LEMA 3.4. *Sejam v_1 e v_2 vetores não nulos do plano. Se v_1 é múltiplo de v_2 , então as normalizações de v_1 e v_2 são iguais, a menos de sinal. Em particular, se $a > 0$, as duas normalizações são iguais.*

DEMONSTRAÇÃO. Digamos que $v_1 = av_2$, para algum número real a . Então

$$\|v_1\| = |a|\|v_2\|,$$

de modo que

$$\frac{v_1}{\|v_1\|} = \frac{av_2}{|a|\|v_2\|} = \frac{a}{|a|} \frac{v_2}{\|v_2\|} = \pm \frac{v_2}{\|v_2\|},$$

já que a e $|a|$ diferem apenas pelo sinal. Além disso, como $|a| = a$, quando $a > 0$, as normalizações coincidirão neste caso. \square

O segundo lema é consequência do primeiro e nos permite reescrever a recursão u_i definida em (42) de uma maneira mais adequada aos cálculos que faremos na demonstração da convergência do método da potência.

LEMA 3.5. *Se T é um operador linear autoadjunto cujos autovalores não são nulos e u_0 é um vetor do plano, então*

$$u_i = \frac{T^i(u_0)}{\|T^i(u_0)\|}$$

DEMONSTRAÇÃO. Denotando por n_j o inverso da norma $T(u_j)$, temos

$$u_i = n_{i-1}T(u_{i-1}) \quad \text{e} \quad u_{i-1} = n_{i-2}T(u_{i-2}).$$

Substituindo a segunda igualdade na primeira, e lembrando que T é linear,

$$u_i = n_{i-1}n_{i-2}T^2(u_{i-2}).$$

Mas, $u_{i-2} = n_{i-3}T(u_{i-3})$, de modo que

$$u_i = n_{i-1}n_{i-2}T^2(n_{i-3}T(u_{i-3})) = n_{i-1}n_{i-2}n_{i-3}T^3(u_{i-3}).$$

Repetindo este argumento tantas vezes quanto necessário, obtemos finalmente

$$u_i = n_{i-1}n_{i-2} \cdots n_0 T^i(u_0).$$

Portanto, como $n_j > 0$, temos pelo lema, que u_i é igual à normalização de $T^i(u_0)$. \square

TEOREMA 3.6. *Seja T um operador autoadjunto do plano cujos autovalores distintos λ_1 e λ_2 satisfazem $|\lambda_1| > |\lambda_2|$. Se u_0 é um vetor unitário que não é um autovetor associado a λ_2 , então a sequência u_0, u_1, \dots definida em (42) tende a um autovetor de T associado a λ_1 .*

DEMONSTRAÇÃO. Sejam, agora, w_1 e w_2 autovetores unitários de T associados, respectivamente, aos autovalores λ_1 e λ_2 . Como estamos supondo que $\lambda_1 \neq \lambda_2$, os vetores w_1 e w_2 são ortogonais. Logo $\beta = \{w_1, w_2\}$ é uma base ortonormal do plano e podemos escrever

$$u_0 = a_1 w_1 + a_2 w_2,$$

com $a_1, a_2 \in \mathbb{R}$. Além disso, sabemos que $a_1 \neq 0$, porque estamos supondo que u_0 não é múltiplo de w_2 . Porém,

$$T^i(w_1) = T^{i-1}(T(w_1)) = T^{i-1}(\lambda_1 w_1) = \lambda_1 T^{i-1}(w_1)$$

pois w_1 é um autovetor associado a λ_1 . Repetindo o mesmo argumento

$$T^i(w_1) = \lambda_1 T^{i-1}(w_1) = \lambda_1 T^{i-2}(T(w_1)) = \lambda_1 T^{i-2}(\lambda_1 w_1) = \lambda_1^2 T^{i-2}(w_1).$$

Continuando assim, chegamos à conclusão que

$$(43) \quad T^i(w_1) = \lambda_1^i w_1.$$

Analogamente, podemos provar que

$$(44) \quad T^i(w_2) = \lambda_2^i w_2.$$

Mas, substituindo (43) e (44) em

$$T^i(u_0) = a_1 T^i(w_1) + a_2 T^i(w_2),$$

obtemos

$$T^i(u_0) = \lambda_1^i a_1 w_1 + \lambda_2^i a_2 w_2.$$

Dividindo a expressão anterior por λ_1^i ,

$$\frac{T^i(u_0)}{\lambda_1^i} = a_1 w_1 + \left(\frac{\lambda_2}{\lambda_1} \right)^i a_2 w_2.$$

Como estamos supondo que $|\lambda_1| > |\lambda_2|$, a fração λ_2/λ_1 tem valor absoluto menor do que 1. Isto significa que $|\lambda_2/\lambda_1|^i$ diminui à medida que i aumenta. Em particular,

$$\lim_{i \rightarrow +\infty} \left(\frac{\lambda_2}{\lambda_1} \right)^i = 0.$$

Logo,

$$\lim_{i \rightarrow +\infty} \left(\frac{T^i(u_0)}{\lambda_1^i} \right) = \lim_{i \rightarrow +\infty} \left(a_1 w_1 + \left(\frac{\lambda_2}{\lambda_1} \right)^i a_2 w_2 \right) = a_1 w_1;$$

donde

$$\lim_{i \rightarrow +\infty} \left\| \frac{T^i(u_0)}{\lambda_1^i} \right\| = \left\| \lim_{i \rightarrow +\infty} \left(\frac{T^i(u_0)}{\lambda_1^i} \right) \right\| = \|a_1 w_1\| = |a_1| \|w_1\|.$$

Portanto,

$$\lim_{i \rightarrow +\infty} \frac{\frac{T^i(u_0)}{\lambda_1^i}}{\left\| \frac{T^i(u_0)}{\lambda_1^i} \right\|} = \frac{a_1 w_1}{|a_1| \|w_1\|} = \pm \frac{w_1}{\|w_1\|}.$$

Finalmente, pelo lema 3.5

$$\lim_{i \rightarrow +\infty} u_i = \lim_{i \rightarrow +\infty} \frac{T^i(u_0)}{\|T^i(u_0)\|} = \lim_{i \rightarrow +\infty} \frac{\frac{T^i(u_0)}{\lambda_1^i}}{\left\| \frac{T^i(u_0)}{\lambda_1^i} \right\|} = \pm \frac{w_1}{\|w_1\|},$$

como queríamos mostrar. \square

O teorema garante que, sob certas hipóteses, o método da potência nos permite calcular uma aproximação para um dos autovetores de um operador. Tendo encontrado o autovetor, podemos obter o autovalor a ele associado sem a necessidade de achar as raízes do polinômio característico. Para entender o porquê disto, lembre-se que se w é autovetor de um operador T associado ao autovalor λ , então $T(w) = \lambda w$. Portanto, se $w = (a_1, a_2)$ e $T(w) = (b_1, b_2)$, então de $(b_1, b_2) = (\lambda a_1, \lambda a_2)$ obtemos

$$\lambda = \frac{b_i}{a_i}$$

quando $a_i \neq 0$, o que tem que ocorrer para uma das coordenadas de w porque, sendo um autovetor, w não pode ser nula. Por exemplo, aplicando o método da potência ao operador T definido em (45) a partir do vetor $u_0 = (1, 1)$, obtemos ao final de dez iterações a aproximação

$$u_{10} = (0.923879, -0.382683)$$

para o autovetor de seu autovalor λ de maior módulo. Logo,

$$T(u_{10}) \approx \lambda u_{10}.$$

Mas,

$$T(u_{10}) = (-7.839372, 3.247174),$$

onde

$$(-7.839372, 3.247174) \approx \lambda(0.923879, -0.382683).$$

Obtemos, assim, da primeira coordenada

$$\lambda \approx \frac{-7.839372}{0.923879} \approx -8.485279$$

e da segunda coordenada

$$\lambda \approx \frac{3.247174}{-0.382683} \approx -8.485284$$

Naturalmente aumentando a quantidade de iterações obteremos uma aproximação melhor.

Ao contrário do procedimento analítico, que requer que calculemos os autovalores do operador e só então seus autovetores, o método da potência nos dá diretamente o autovetor associado ao autovalor de maior módulo. Na verdade, *não é viável usar aproximações das raízes do polinômio característico no cálculo dos autovetores*, porque o sistema linear resultante *não seria indeterminado*. Por exemplo, o operador linear cuja matriz é

$$(45) \quad T(x, y) = (x + 3y, 3x + 7y)$$

tem polinômio característico igual a $t^2 - 8t - 2$, de modo que seus autovalores são $4 - 3\sqrt{2}$ e $3\sqrt{2} + 4$. Suponhamos que desejamos calcular os autovetores associados ao maior autovalor, mas tudo o que conhecemos é a aproximação

$$3\sqrt{2} + 4 \approx 8.242640687119286$$

Para achar os autovetores, precisaríamos resolver o sistema linear

$$\begin{aligned} -7.242640687119286x + 3y &= 0 \\ 3x - 1.242640687119285y &= 0. \end{aligned}$$

No entanto, este sistema *não* é indeterminado, pois

$$\det \begin{bmatrix} -7.242640687119286 & 3 \\ 3 & -1.242640687119285 \end{bmatrix} = 3.552713678800501 \cdot 10^{-15}$$

apesar de muito pequeno, é diferente de zero.

Antes que você se pergunte porque alguém escolheria usar aproximações no cálculo dos autovetores, em vez de executá-los de maneira exata, aqui vão duas razões. A primeira é que o mesmo método funciona para operadores definidos por matrizes de tamanho $n \times n$. Nestes casos o polinômio característico terá grau n , mas não há fórmulas para achar raízes de polinômios de grau maior que 4. Assim, a única saída é calcular aproximações. A segunda razão é que matrizes oriundas de medidas são intrinsecamente dadas por valores aproximados, o que torna impossível calcular as raízes de maneira exata, mesmo que a matriz seja 2×2 .

Finalmente, é importante lembrar que o teorema de convergência do método da potência requer que os autovalores não tenham ambos o mesmo módulo. Caso isto não aconteça, pode não haver convergência. Por exemplo, aplicando o método da potência à reflexão $R(x, y) = (x, -y)$ a partir de $u_0 = (0, 1)$, obtemos a sequência de vetores

$$u_i = \begin{cases} (0, 1) & \text{quando } i \text{ é par} \\ (0, -1) & \text{quando } i \text{ é ímpar}, \end{cases}$$

de modo que não há convergência para nenhum vetor. Para o caso de você achar que este exemplo não é justo porque, afinal de contas, $(0, 1)$ já é um autovetor do operador R , aqui vai outro exemplo. Calculando cinco iterações do método da potência aplicado ao operador cuja matriz na base canônica é

$$\begin{bmatrix} -\frac{6}{5} & -\frac{8}{5} \\ -\frac{8}{5} & \frac{6}{5} \end{bmatrix}$$

partindo de $u_0 = (1, 1)$ obtemos

$$u_i = \begin{cases} (-0.9899494936611664, -0.1414213562373095) & \text{quando } i = 1, 3, 5 \\ (0.7071067811865475, 0.7071067811865474) & \text{quando } i = 2, 4, 6. \end{cases}$$

Mesmo aumentando muito a quantidade de iterações, não encontramos indícios de convergência. Por exemplo,

$$u_{99} = (-0.9899494936, -0.1414213562) \quad \text{e} \quad u_{100} = (0.7071067811, 0.70710678).$$

Exercícios

1. Sejam A e B matrizes quadradas 2×2 . Prove as seguintes propriedades da transposta de uma matriz:

- (a) $(A^t)^t = A$;
- (b) $(A + B)^t = A^t + B^t$;
- (c) $(AB)^t = B^t A^t$.

Note a troca da ordem em que as matrizes são multiplicadas no item (c).

2. Determine os autovalores e autovetores das seguintes matrizes:

$$(a) \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad (b) \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad (c) \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad (d) \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

3. Calcule os autovalores e autovetores de cada uma das matrizes abaixo e determine se são diagonalizáveis:

$$A = \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{e} \quad C = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}.$$

4. Considere o operador linear T de \mathbb{R}^2 dado por $T(x, y) = (5x - y, -x + 5y)$.

- (a) Encontre os vértices de um retângulo que tenha a origem como um dos seus vértices e que seja levado por T num quadrado de lado 12.
- (b) Encontre a imagem por T do triângulo de vértices $A = (0, 0)$, $B = (2, 2)$ e $C = (1, -1)$.

5. Determine a transformação linear que descreve o movimento rígido que leva o segmento de extremos $A = (-6, 2)$ e $B = (-1, 2)$ no segmento de extremos $C = (-2, 6)$ e $D = (1, 2)$, respectivamente. Prove que esta transformação é uma rotação e calcule seu ângulo.

6. Seja T um operador autoadjunto de \mathbb{R}^2 . Sabe-se que:

- (a) os autovalores de T são 2 e 3;
- (b) $(1, 1)$ é um autovetor de T associado ao autovalor 2.

Determine a matriz de T na base canônica.

7. Prove que se P é a matriz de uma projeção do plano em uma reta então P é simétrica e $P^2 = P$.
8. Seja P a matriz de um operador linear do plano. Prove que se P é simétrica e $P^2 = P$, então o operador que corresponde a P é uma projeção do plano em uma reta.
9. Mostre que um operador linear autoadjunto do plano cujos autovalores são 1 e -1 é, necessariamente, uma reflexão.
10. Seja T um operador autoadjunto do plano que tem dois autovalores distintos, um dos quais é λ . Mostre que quaisquer dois autovetores de T associados a λ são colineares.
11. Dizemos que dois operadores autoadjuntos T_1 e T_2 comutam se

$$T_1(T_2(v)) = T_2(T_1(v)) \quad \text{qualquer que seja o vetor } v \in \mathbb{R}^2.$$

Use a fórmula (39) e o exercício 7 da página 33 para mostrar que

$$(T_1)_\beta(T_2)_\beta = (T_2)_\beta(T_1)_\beta$$

qualquer que seja a base ortonormal β do plano.

12. Sejam T_1 e T_2 dois operadores autoadjuntos e seja β uma base ortonormal que diagonaliza tanto T_1 , quanto T_2 .
 - (a) Mostre que $(T_1)_\beta$ e $(T_2)_\beta$ comutam.
 - (b) Use mudança de base para relacionar $(T_1)_\beta$ a $(T_1)_\varepsilon$ e $(T_2)_\beta$ a $(T_2)_\varepsilon$.
 - (c) Use (a) e (b) para provar que $(T_1)_\varepsilon$ e $(T_2)_\varepsilon$ também comutam.
13. Mostre que os operadores autoadjuntos do plano definidos pelas regras

$$T_1(x, y) = (x + 2y, 2x + y) \quad \text{e} \quad T_2(x, y) = (3x + 2y, 2x + y)$$
 não podem ser diagonalizados por uma mesma base ortonormal.
14. Sejam T_1 e T_2 dois operadores autoadjuntos que comutam e suponhamos que T_1 tenha dois autovalores distintos.
 - (a) Mostre que se u_1 é um autovetor de T_1 associado ao autovalor λ_1 , então o mesmo vale para $T_2(u_1)$.
 - (b) Use (a) e o exercício 10 para mostrar que u_1 também é autovetor de T_2 .
 - (c) Use (b) para mostrar que qualquer base ortonormal de autovetores de T_1 também é uma base ortonormal formada por autovetores de T_2 .

Parte 2

Ferramentas

Na segunda parte do curso introduzimos matrizes $n \times n$ e estudamos o método de eliminação, que será usado para resolver muitos dos problemas que surgirão quando generalizarmos para n dimensões tudo o que fizemos em duas dimensões na primeira parte.

CAPÍTULO 3

Matrizes

Como vimos no capítulo 1 as matrizes foram introduzidas por A. Cayley no século XIX como uma maneira prática de efetuar cálculos com transformações lineares. Neste capítulo introduzimos matrizes $m \times n$ e apresentamos os principais conceitos e resultados referentes às matrizes que serão necessários no resto do curso.

1. Definição e exemplos

Além da introduzir a definição geral de matriz, veremos nesta seção vários exemplos de matrizes especiais que serão muito úteis no resto do curso.

1.1. Definição e primeiros exemplos. Uma *matriz* nada mais é do que uma tabela de números com m colunas e n linhas, com m e n dois números inteiros positivos quaisquer. Neste caso dizemos que a matriz tem *tamanho* $m \times n$. Como seria de esperar, as matrizes para as quais $m = n$ são chamadas de *quadradas*; as demais são conhecidas como matrizes *retangulares*. Cada casa da tabela em que a matriz consiste pode ser preenchida com um número real qualquer. Estes números são conhecidos como *entradas* ou *coeficientes* da matriz e a tabela na qual são dispostos é encapsulada por colchetes ou parênteses. Como matrizes são, em última análise, apenas tabelas de números, é natural dizer que duas matrizes são *iguais* quando as tabelas de números correspondentes coincidem; em outras palavras, quando cada casa da tabela é preenchida, nas duas matrizes, com o mesmo valor.

Para não ter que repetir todo o quadro numérico cada vez que nos referimos a uma matriz, vamos designá-las por letras, geralmente maiúsculas. Por exemplo,

$$M = \begin{bmatrix} 1 & 5 & \pi & -5/7 \\ 1/8 & 9 & 8 & \pi/2 \\ 0 & -65 & 0 & 7/\pi \end{bmatrix}$$

é uma matriz com 3×4 , pois tem 3 linhas e 4 colunas. Como ocorre em qualquer tabela, localizamos uma entrada de uma matriz, definindo sua posição em termos da linha e da coluna que ocupa. Por exemplo, na matriz M acima, $\pi/2$ ocupa a posição 2, 4 e -65 a posição 3, 2. Como frases do tipo “o número α ocupa a posição que está na interseção

da linha i com a coluna j da matriz A ” são muito verbosas, vamos abreviá-las escrevendo simplesmente

$$a_{i,j} = \alpha.$$

Note que as entradas da matriz serão denotadas usando, na medida do possível, a letra minúscula correspondente à maiúscula usada para denotar a matriz. Assim, tomando como base a matriz M do exemplo acima mais uma vez, temos

$$m_{1,4} = -5/7 \text{ e } m_{2,2} = 9.$$

Usando esta maneira de representar os coeficientes, podemos escrever a definição de igualdade entre duas matrizes A e B de tamanho $m \times n$ na forma

$$a_{i,j} = b_{i,j} \quad \text{para todo } 1 \leq i \leq m \quad \text{e todo } 1 \leq j \leq n.$$

A *diagonal de uma matriz* A corresponde às posições $a_{i,i}$. Na matriz do exemplo, a diagonal é formada pelas entradas

$$m_{1,1} = 1, m_{2,2} = 9 \text{ e } m_{3,3} = 0.$$

Naturalmente a diagonal de uma matriz só se parece com uma diagonal, no sentido geométrico do termo, quando a matriz é quadrada. Diremos que uma matriz quadrada é *diagonal* se suas únicas entradas não nulas pertencem à sua diagonal. Por exemplo, a matriz

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \pi/2 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} \quad \begin{array}{l} \text{é diagonal, já} \\ \text{não é.} \end{array}$$

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

A mais importante de todas as matrizes diagonais é a matriz *identidade*. Denotada por I , ou I_n quando for necessário deixar claro que se trata de uma matriz $n \times n$, ela tem 1s ao longo da diagonal e zeros em todas as outras posições, como é o caso de

$$I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Por uma questão de economia de espaço, frequentemente denotaremos a matriz identidade cujas entradas na diagonal são d_1, d_2, \dots, d_n por $\text{diag}(d_1, d_2, \dots, d_n)$. Por exemplo, $I_3 = \text{diag}(1, 1, 1)$.

Nem sempre é conveniente definir uma matriz apresentando-a como um quadro de números. Isto ocorre, por exemplo, se a matriz for *esparsa*; isto é, se a maioria de suas entradas forem nulas, como é o caso da matriz identidade. Imagine desenhar uma matriz identidade 100×100 : o quadro numérico é enorme, mas está quase todo ocupado por zeros.

Uma maneira mais econômica de definir tais matrizes consiste em defini-las coeficiente a coeficiente. Por exemplo, a matriz identidade $n \times n$ pode ser definida por

$$I_n[i, j] = \begin{cases} 1 & \text{quando } i = j \\ 0 & \text{quando } i \neq j \end{cases}$$

Note que escolhemos pôr os índices que identificam a posição da entrada entre colchetes, em vez de usar subscritos, para evitar confusão com o n que identifica a dimensão da matriz. Uma matriz ainda mais fácil de definir usando esta terminologia é a matriz nula, denotada simplesmente por 0, porque todos os seus coeficientes são iguais a zero.

1.2. Matrizes especiais. As duas matrizes especiais mais importantes são, provavelmente, a matriz identidade, definida ao final do artigo anterior e a *matriz nula* cujas entradas são todas iguais a zero. Ambas são matrizes esparsas, porque a maioria (ou todas) as suas entradas são nulas. Outra classe importante de matriz esparsa são as matrizes básicas $E_{r,s}$, cujas posições são todas nulas, exceto pela posição r, s , que é igual a 1. Denotando por $e_{i,j}$ a entrada na posição i, j de E_{rs} , temos

$$e_{i,j} = \begin{cases} 1 & \text{quando } i = r \text{ e } j = s \\ 0 & \text{em qualquer outro caso.} \end{cases}$$

Por exemplo,

$$E_{1,3} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{e} \quad E_{3,2} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Uma segunda classe importante de matrizes esparsas são as matrizes *elementares*, nas quais todas as entradas são nulas, exceto a diagonal, cujas entradas são todas iguais a 1 e uma única outra posição, fora da diagonal. A matriz elementar cuja entrada não nula fora da diagonal está na i -ésima linha e j -ésima coluna e é ocupada pelo número real α , será denotada por $L_{rs}(\alpha)$. Denotando por ℓ_{rs} a entrada na posição i, j de $L_{rs}(\alpha)$, temos

$$\ell_{i,j} = \begin{cases} 1 & \text{quando } i = j \\ \alpha & \text{quando } i = r \text{ e } j = s \\ 0 & \text{em qualquer outro caso.} \end{cases}$$

Por exemplo,

$$L_{1,3}(-5) = \begin{bmatrix} 1 & 0 & -5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{e} \quad L_{3,2}(9) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 9 & 1 \end{bmatrix}.$$

As matrizes elementares são exemplos de uma classe mais ampla de matrizes, chamadas de triangulares. Uma matriz U é *triangular superior* se todas as suas entradas abaixo da

diagonal são nulas; isto é, se $U_{i,j} = 0$, sempre que $i < j$. Analogamente, L é uma matriz *triangular inferior* se suas entradas acima da diagonal são nulas, o que equivale a dizer que $L_{i,j} = 0$ quando $i > j$. Nos exemplos abaixo, U é triangular superior, ao passo que L é triangular inferior

$$U = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 6 & -3 & 0 \\ 0 & 0 & 8 & -9 \\ 0 & 0 & 0 & 14 \end{bmatrix} \quad \text{e} \quad L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & -6 & 0 & 0 \\ -9 & 4 & 8 & 0 \\ 11 & -2 & 0 & 14 \end{bmatrix}.$$

Como vimos no capítulo 2, uma matriz quadrada A , de tamanho $n \times n$, é *simétrica* quando suas entradas satisfazem $a_{i,j} = a_{j,i}$, para todo $1 \leq i < j \leq n$. Por outro lado, A é *antissimétrica* quando $a_{i,j} = -a_{j,i}$, para todo $1 \leq i < j \leq n$. Assim, entre as matrizes

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 7 \\ 3 & 7 & 9 \end{bmatrix} \quad \text{e} \quad \begin{bmatrix} 1 & 2 & 3 \\ -2 & 5 & -7 \\ -3 & 7 & 9 \end{bmatrix}$$

a primeira é simétrica e a segunda é antissimétrica. Teremos muito mais a dizer sobre as matrizes simétricas no capítulo 8. Embora as matrizes antissimétricas não tenham um papel tão preponderante neste livro, elas desempenham papel chave na geometria simplética, que é a tradução geométrica da mecânica clássica.

As matrizes de Vandermonde, que desempenham papel essencial nos problemas de interpolação, são facilmente construídas pela regra que define seus coeficientes. Dados n números reais $\alpha_1, \dots, \alpha_n$, a entrada i, j da matriz de Vandermonde $V = V(\alpha_1, \dots, \alpha_n)$, é definida pela regra

$$v_{i,j} = \alpha_i^{j-1}.$$

Quando $n = 3$ isto nos dá

$$\begin{bmatrix} 1 & \alpha_1 & \alpha_1^2 \\ 1 & \alpha_2 & \alpha_2^2 \\ 1 & \alpha_3 & \alpha_3^2 \end{bmatrix}$$

Todas as matrizes que construímos até agora são quadradas, mas as matrizes retangulares não serão menos importantes neste livro. Destas, há duas para as quais convém chamar a atenção desde já: as matrizes linha e as matrizes colunas. As primeiras têm uma única linha; as segundas, uma única coluna. No capítulo 1 usamos estas últimas para representar os vetores do plano de forma matricial.

2. Operações com matrizes

Há quatro operações com matrizes que precisamos considerar: a adição, a multiplicação de uma matriz por um número real, a multiplicação de duas matrizes e a transposição de matrizes. As três primeiras operações são binárias, porque combinam dois ingredientes para produzir uma nova matriz; no caso da adição e multiplicação de matrizes, estes ingredientes são duas matrizes, ao passo que na multiplicação por escalar, os ingredientes são um número real e uma matriz. Por outro lado, a transposição é uma operação unária, porque tem como único ingrediente uma matriz.

2.1. Adição e multiplicação por escalar. A adição e a multiplicação por escalar aplicam-se a quaisquer matrizes e são definidas coeficiente a coeficiente. Mais precisamente, se A e B são duas matrizes $m \times n$ e λ é um número real, então a posição i, j de $S = A + B$ é definida por

$$S_{i,j} = a_{i,j} + b_{i,j}$$

ao passo que a mesma posição em $E = \lambda A$ é definida por

$$E_{i,j} = \lambda a_{i,j}.$$

Note que só podemos somar duas matrizes quando ambas têm o mesmo tamanho. Por exemplo, se $\lambda = 3$,

$$A = \begin{bmatrix} 0 & 1 & -2 \\ 3 & -1 & 7 \\ 6 & -8 & 4 \end{bmatrix} \quad \text{e} \quad B = \begin{bmatrix} 9 & -1 & 2 \\ 5 & 6 & 3 \\ -3 & 0 & -8 \end{bmatrix},$$

então

$$A + B = \begin{bmatrix} 9 & 0 & 0 \\ 8 & 5 & 10 \\ 3 & -8 & -4 \end{bmatrix} \quad \text{e} \quad \lambda A = 3A = \begin{bmatrix} 0 & 3 & -6 \\ 9 & -3 & 21 \\ 18 & -24 & 12 \end{bmatrix}.$$

A adição de matrizes satisfaz essencialmente as mesmas propriedades que a adição de números reais e podem ser facilmente provadas a partir destas últimas. Se A , B e C são matrizes $m \times n$, então as propriedades que precisamos analisar são as seguintes:

associatividade: $A + (B + C) = (A + B) + C$;

comutatividade: $A + B = B + A$;

elemento neutro: $A + 0 = A$;

em que 0 representa a matriz $m \times n$ cujas entradas são todas nulas. Vamos provar a primeira propriedade, as outras duas são mais fáceis e ficarão por sua conta. Para começar, lembre-se que duas matrizes são iguais se têm as mesmas entradas. Portanto, para mostrar que

$S = A + (B + C)$ é igual a $S' = (A + B) + C$ basta mostrar que $S_{i,j} = S'_{i,j}$, quaisquer que sejam $1 \leq i \leq m$ e $1 \leq j \leq n$. Contudo, pela definição de soma de matrizes

$$S_{i,j} = a_{i,j} + (b_{i,j} + c_{i,j}) \quad \text{e} \quad S'_{i,j} = (a_{i,j} + b_{i,j}) + c_{i,j},$$

que são iguais porque a soma de números reais é associativa. Em outras palavras, a demonstração da associatividade da adição de matrizes consiste em reduzir o problema à associativa da soma de suas entradas.

Apesar de ser a menos prestigiada das três propriedades listadas acima, a associatividade é uma das mais úteis. Para poder apreciá-la devidamente, lembre-se que a adição de matrizes é uma operação binária; isto é, só podemos somar duas matrizes de cada vez. Assim, a única maneira de somar três matrizes é calcular a soma de duas delas e, então, somar este resultado à terceira. O problema é que há duas maneiras de fazer isto. Podemos calcular primeiro $A+B$ e somar o resultado a C ou podemos calcular $B+C$ e somar A com esta matriz. A associatividade nos diz que, qualquer das duas escolhas produz o mesmo resultado. É graças a esta propriedade que podemos escrever $A_1 + A_2 + A_3 + A_4 + A_5$ para a soma de três matrizes, em vez de

$$((A_1 + (A_2 + A_3) + A_4) + A_5).$$

 Convém lembrar que o universo está cheio de operações que não são associativas. Digamos, por exemplo, que sua calculadora representa números em notação científica, na forma

$$a_0.a_1a_2a_3a_4a_5a_6a_7a_8 \times 10^n$$

em que $0 \leq a_0, \dots, a_8 \leq 9$ e $-99 \leq n \leq 99$ são inteiros. Se

$$\alpha = -10^{30}, \quad \beta = 10^{30} \quad \text{e} \quad \gamma = 1$$

então, como esperado,

$$(\alpha + \beta) + \gamma = \gamma = 1.$$

Porém

$$\beta + \gamma = 1 + 10^{30}$$

será arredondado pela calculadora para 10^{30} , porque não há espaço no mostrador para escrever simultaneamente o 1 que fica na casa das unidades e o 1 situado 30 casas decimais acima e que corresponde a 10^{30} . Por causa disto, a calculadora retornará 0 como resultado de $\alpha + (\beta + \gamma)$, como você mesmo pode verificar.

Passando às propriedades da multiplicação de uma matriz por um escalar, temos que, dadas duas matrizes A e B de tamanho $m \times n$ e dois números reais λ e μ ,

- $0 \cdot A = 0$;
- $1 \cdot A = A$;
- $\lambda(A + B) = \lambda A + \lambda B$;
- $(\lambda + \mu)A = \lambda A + \mu A$;
- $\lambda(\mu A) = \lambda(\mu A)$;

Todas estas propriedades podem ser provadas de maneira análoga a que usamos para provar a associatividade da adição, por isso ficarão por sua conta. A matriz obtida multiplicando -1 pela matriz A é geralmente denotada simplesmente por $-A$, pois a posição i, j de $-A$ é igual a $-a_{i,j}$. Em particular, $A + (-A) = 0$ é a matriz nula de mesmo tamanho que A .

Como aplicação do que aprendemos, podemos escrever uma matriz A de tamanho $n \times m$ como uma soma de escalares multiplicados pelas matrizes básicas $E_{i, \cdot}$. Mais precisamente,

$$(46) \quad A = (a_{1,1}E_{1,1} + \cdots a_{1,n}E_{1,n}) + \cdots + (a_{m,1}E_{m,1} + \cdots a_{m,n}E_{m,n}).$$

Por sua vez, as matrizes elementares podem ser escritas de maneira ainda muito simples como

$$L_{i,j}(\alpha) = I + \alpha E_{i,j},$$

em que I e $E_{i,j}$ têm o mesmo tamanho que $L_{i,j}(\alpha)$.

2.2. Multiplicação de matrizes. Nossa objetivo neste artigo é adaptar a regra que descobrimos para a multiplicação de matrizes 2×2 para o caso geral em que as matrizes não são nem mesmo quadradas. A maneira usual de descrevê-la recorre a uma fórmula geral, cheia de coeficientes. Mas há uma maneira mais civilizada de expressá-la. Começamos com o caso em que uma das matrizes é uma linha e a outra uma coluna; isto é,

$$L = \begin{bmatrix} \ell_{1,1} & \ell_{1,2} \end{bmatrix} \text{ e } C = \begin{bmatrix} c_{1,1} \\ c_{2,1} \end{bmatrix}$$

A regra deduzida no artigo 3.1 do capítulo 1 sugere que deveríamos definir o produto $L \cdot C$ como sendo a matriz 1×1 cuja única entrada é $\ell_{1,1}c_{1,1} + \ell_{1,2}c_{2,1}$. Podemos considerar isto como uma matriz 1×1 ou como um número real, isto é, um escalar. Em geral, se

$$L = \begin{bmatrix} \ell_{1,1} & \cdots & \ell_{1,n} \end{bmatrix} \text{ e } C = \begin{bmatrix} c_{1,1} \\ \vdots \\ c_{n,1} \end{bmatrix}$$

copiamos a definição acima, escrevendo,

$$(47) \quad L \cdot C = \ell_{1,1}c_{1,1} + \cdots + \ell_{1,n}c_{n,1}.$$

Por exemplo, quando

$$L = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \text{ e } C = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$$

obtemos

$$L \cdot C = 1 \cdot 4 + 2 \cdot 5 + 3 \cdot 6 = 32.$$

Note que escolhemos L como tendo n colunas e C como tendo n linhas, do contrário sobrariam coeficientes em L ou C quando fôssemos calcular o somatório que define $L \cdot C$. Em outras palavras,

só faz sentido multiplicar uma matriz linha $1 \times n$ por uma matriz coluna $m \times 1$ quando $n = m$.

A propósito, as matrizes $1 \times n$ são conhecidas como *matrizes linha* e as $r \times 1$ como *matrizes coluna*.

Para estender isto a matrizes A e B de tamanhos $m \times n$ e $r \times s$, respectivamente, consideraremos cada linha de A como sendo uma matriz $1 \times n$ e cada coluna de B como sendo uma matriz $r \times 1$. A primeira coisa a notar é que, para que seja possível multiplicar uma linha de A por uma coluna de B devemos ter que $n = r$. Sob esta condição, definiremos a entrada i, j da matriz produto AB como sendo o escalar que resulta do produto da i -ésima linha de A pela j -ésima coluna de B . Para escrever uma fórmula explícita é conveniente ter uma notação para linhas e colunas de uma matriz. Utilizando a terminologia do SCILAB escreveremos $A_{i,*}$ para denotar a i -ésima linha e $A_{*,i}$ para denotar a i -ésima coluna da matriz A . Com isto, a fórmula que define a matriz produto $M = AB$ é

$$(48) \quad m_{i,j} = A_{i,*} \cdot B_{*,j} = a_{i,1}b_{1,j} + \cdots + a_{i,n}b_{n,j}.$$

Note que i percorre os índices das *linhas de A*, ao passo que j percorre os índices das *colunas de B*. Como AB tem uma entrada para cada i e cada j , sua dimensão será $m \times s$. Temos, assim, que

o produto de uma matriz $m \times n$ por uma matriz $r \times s$ só existe se $n = r$; neste caso o produto será uma matriz $m \times s$.

Podemos deduzir diretamente de (48) duas fórmulas muito úteis para as linhas e as colunas de AB . No caso das linhas,

$$(49) \quad (AB)_{i,*} = A_{i,*}B,$$

já no caso das colunas,

$$(50) \quad (AB)_{*,j} = A_{*,1}B_{1,j} + A_{*,2}B_{2,j} + \cdots + A_{*,n}B_{n,j}.$$

Começaremos aplicando a fórmula (48) para calcular o produto de duas matrizes básicas. Seja $E_{r,s}$ e $E_{u,v}$ duas matrizes básicas de tamanhos $m \times n$ e $n \times k$, respectivamente, e seja $M = E_{r,s}E_{u,v}$ seu produto. Mas, denotando as entradas de $E_{r,s}$ por $e_{i,j}$ e de $E_{u,v}$ por $e'_{i,j}$, temos que

$$(51) \quad m_{i,j} = e_{i,1}e'_{1,j} + e_{i,2}e'_{2,j} + \cdots + e_{i,n}e'_{n,j}.$$

Mas as matrizes básicas têm apenas uma entrada não nula. Logo, $m_{i,j}$ só pode ser diferente de zero se $e_{i,\ell} \neq 0$ e $e'_{\ell,j} \neq 0$, para algum $1 \leq \ell \leq n$. Porém, para que isto ocorra é

necessário que $i = r$, $\ell = s = u$ e $j = v$. Portanto,

$$(52) \quad E_{r,s}E_{u,v} = \begin{cases} E_{r,v} & \text{se } s = v \\ 0 & \text{em qualquer outro caso.} \end{cases}$$

Também é fácil usar a fórmula (48) para provar que a matriz identidade é o elemento neutro da multiplicação de matrizes. Sejam A uma matriz $n \times n$ e I a matriz identidade de mesmo tamanho; nosso objetivo é mostrar que

$$A \cdot I = I \cdot A = A.$$

Pela fórmula (48), a entrada i, j de $A \cdot I$ é igual a

$$A_{i,*} \cdot I_{*,j}.$$

Mas $I_{*,j}$ tem apenas uma entrada não nula, que fica na posição j, j . Portanto, pela fórmula (47),

$$A_{i,*} \cdot I_{*,j} = a_{i,j}.$$

Logo, a entrada i, j de $A \cdot I$ é igual a $a_{i,j}$, que é a entrada correspondente na matriz A , provando, assim, a igualdade destas duas matrizes. Um argumento análogo mostra que $I \cdot A = A$. Portanto, a matriz identidade é o *elemento neutro* da multiplicação de matrizes.

2.3. Propriedades da multiplicação de matrizes. Vimos no artigo anterior que a matriz identidade funciona como elemento neutro para a multiplicação de matrizes. Outra propriedade fácil de provar é a *distributiva*, segundo a qual

$$A(B + C) = AB + AC$$

em que A é uma matriz $m \times n$ e B e C são matrizes $n \times r$. Como a matriz resultante destes produtos, de um lado e do outro da igualdade, têm tamanho $m \times r$, basta provarmos que se $S = A(B + C)$ e $S' = AB + AC$, então $S_{i,j} = S'_{i,j}$, para todo $1 \leq i \leq m$ e $1 \leq j \leq r$. Mas, pela fórmula (48) de multiplicação de matrizes, as entradas i, j de AB e de AC são iguais a, respectivamente,

$$a_{i,1}b_{1,j} + \cdots + a_{i,n}b_{n,j} \quad \text{e} \quad a_{i,1}c_{1,j} + \cdots + a_{i,n}c_{n,j}.$$

Logo, como a soma de matrizes é feita coordenada a coordenada,

$$S'_{i,j} = (a_{i,1}b_{1,j} + \cdots + a_{i,n}b_{n,j}) + (a_{i,1}c_{1,j} + \cdots + a_{i,n}c_{n,j});$$

que podemos reagrupar como

$$S'_{i,j} = (a_{i,1}b_{1,j} + a_{i,1}c_{1,j}) + \cdots + (a_{i,n}b_{n,j} + a_{i,n}c_{n,j}).$$

Pondo as entradas de A em evidência para cada soma de duas parcelas,

$$(53) \quad S'_{i,j} = a_{i,1}(b_{1,j} + c_{1,j}) + \cdots + a_{i,n}(b_{n,j} + c_{n,j}).$$

Mas,

$$A_{i,*} = [a_{i,1} \quad \cdots \quad a_{i,n}]$$

é a i -ésima linha de A e

$$(B + C)_{*,j} = \begin{bmatrix} B_{1,j} + L_{1,j} \\ \vdots \\ B_{n,j} + C_{n,j} \end{bmatrix}$$

é a j -ésima coluna de $B + C$, de modo que (53) pode ser reescrita na forma

$$S'_{i,j} = A_{i,*} \cdot (B + C)_{*,j}.$$

Como o lado direito desta última equação é igual a $S_{i,j}$, concluímos que

$$A(B + C) = S = S' = AB + AC$$

como queríamos mostrar.

Como aplicação desta propriedade, vamos calcular o produto entre duas matrizes elementares. Usando a representação das matrizes elementares introduzida no artigo 2.1, podemos escrever

$$L_{r,s}(\alpha) = I + \alpha E_{r,s} \quad \text{e} \quad L_{u,v}(\beta) = I + \beta E_{u,v}.$$

Logo,

$$L_{r,s}(\alpha) \cdot L_{u,v}(\beta) = (I + \alpha E_{r,s})(I + \beta E_{u,v}).$$

Mas, pela propriedade distributiva,

$$(I + \alpha E_{r,s})(I + \beta E_{u,v}) = I + \alpha E_{r,s} \cdot I + I \cdot \beta E_{u,v} + \alpha \beta E_{r,s} E_{u,v}.$$

Logo,

$$L_{r,s}(\alpha) \cdot L_{u,v}(\beta) = I + \alpha E_{r,s} + \beta E_{u,v} + \alpha \beta E_{r,s} E_{u,v}.$$

Portanto, por (52),

$$(54) \quad L_{r,s}(\alpha) \cdot L_{u,v}(\beta) = \begin{cases} I + \alpha E_{r,s} + \beta E_{u,v} & \text{se } s \neq u \\ I + \alpha E_{r,s} + \beta E_{s,v} + \alpha \beta E_{r,v} & \text{se } s = u. \end{cases}$$

Há duas outras propriedades importantes da multiplicação de matrizes que ainda não consideramos: a associatividade e a comutatividade. Como já vimos no artigo 3.1 do capítulo 1, a multiplicação de matrizes não é comutativa nem mesmo quando as matrizes têm tamanho 2×2 . Por outro lado, a multiplicação de matrizes é uma operação associativa; mais precisamente,

$$(55) \quad (AB)C = A(BC),$$

em que, desta vez, A , B e C representam matrizes de tamanhos $m \times n$, $n \times r$ e $r \times s$, respectivamente. Para mostrar que (55) é válida, basta provarmos que as entradas correspondentes a uma mesma posição são iguais nas duas matrizes. Mas, pela fórmula (48) da multiplicação de matrizes, a entrada i, j de $(AB)C$ é igual a

$$(AB)_{i,*} L_{*,j},$$

ao passo que a i -ésima linha da matriz AB é a matriz

$$(AB)_{i,*} = \begin{bmatrix} A_{i,*}B_{*,1} & A_{i,*}B_{*,2} & \cdots & A_{i,*}B_{*,r} \end{bmatrix}$$

Combinando as duas últimas equações, vemos que a entrada i, j de $(AB)C$ é

$$\begin{bmatrix} A_{i,*}B_{*,1} & A_{i,*}B_{*,2} & \cdots & A_{i,*}B_{*,r} \end{bmatrix} \begin{bmatrix} L_{1,j} \\ \vdots \\ L_{r,j} \end{bmatrix}.$$

Efetuando a multiplicação obtemos

$$((AB)C)_{i,j} = A_{i,*}B_{*,1} \cdot L_{1,j} + A_{i,*}B_{*,2} \cdot L_{2,j} + \cdots + A_{i,*}B_{*,r} \cdot L_{r,j}.$$

Como já sabemos que a multiplicação se distribui sobre a soma, esta última equação mostra que a entrada i, j de $(AB)C$ é

$$A_{i,*}(B_{*,1} \cdot L_{1,j} + B_{*,2} \cdot L_{2,j} + \cdots + B_{*,r} \cdot L_{r,j}).$$

Contudo, pela fórmula (50),

$$(BC)_{*,j} = B_{*,1}L_{1,j} + B_{*,2}L_{2,j} + \cdots + B_{*,n}L_{n,j};$$

de modo que a entrada i, j de $(AB)C$ pode ser escrita na forma

$$A_{i,*}(BC)_{*,j}$$

que é igual à entrada i, j de $A(BC)$, como queríamos provar.

2.4. Transposta. Encerraremos esta seção definindo a transposição de matrizes e analisando como se comporta relativamente às demais operações. A *transposta* A^t de uma matriz A é a matriz obtida trocando-se as linhas pelas colunas de A , de modo que

$$(A^t)_{i,*} = A_{*,i};$$

isto é, a entrada i, j de A^t é igual a $a_{j,i}$. Naturalmente, se a matriz A tem tamanho $m \times n$, sua transposta será $n \times m$; afinal, linhas viraram colunas e vice-versa. Naturalmente, a transposta da transposta é a matriz original:

$$(A^t)^t = A.$$

O comportamento da adição e da multiplicação por escalar relativamente à transposição é muito simples. Se A e B forem matrizes de mesmo tamanho, então

$$(A + B)^t = A^t + B^t \text{ e } (\lambda \cdot A)^t = \lambda \cdot A^t;$$

qualquer que seja o escalar λ , como você pode constatar facilmente a partir das definições. Já o comportamento da multiplicação relativamente à transposição é um pouco mais sutil. Suponhamos que A é uma matriz $m \times n$ e B uma matriz $n \times s$. Como o número de linhas de A coincide com o de colunas de B , o produto AB existe e é uma matriz $m \times s$. Portanto, a transposta $(AB)^t$ será uma matriz $s \times m$. Ao contrário do que você possa esperar, esta matriz *não* pode, em geral, ser igual a $A^t \cdot B^t$. Na verdade, como A^t tem tamanho $n \times m$ e

B^t tamanho $s \times n$, o produto $A^t \cdot B^t$ sequer estará definido quando $m \neq s$. Curiosamente, os tamanhos de A^t e B^t nos permitem calcular $B^t A^t$ e este produto de fato coincide com $(AB)^t$. Para comprovar isto, lembre-se que, por definição,

$$(A^t)_{*,j} = A_{j,*} \text{ e } (B^t)_{i,*} = B_{*,i},$$

ao passo que a entrada i, j de $(AB)^t$ é igual à entrada j, i de AB . Mas, pela fórmula (48) da multiplicação de matrizes

$$A_{j,*} \cdot B_{*,i} = (A^t)_{*,j} \cdot (B^t)_{i,*};$$

que é a entrada i, j de $(A^t)(B^t)$, provando, assim, a igualdade desejada. Resumindo, a transposição satisfaz as seguintes propriedades:

$$(A + B)^t = A^t + B^t, \quad (\lambda A)^t = \lambda A^t \quad \text{e} \quad (AB)^t = B^t A^t,$$

quaisquer que o escalar $\lambda \in \mathbb{R}$ e as matriz A e B de tamanho $m \times n$ e C de tamanho $n \times r$.

2.5. Matriz inversa. Uma matriz quadrada A é *inversível* quando existe uma matriz B tal que

$$AB = BA = I,$$

em que I é a matriz identidade. Note que A , B e I são matrizes quadradas de mesmo tamanho. Neste caso, dizemos que B é a *matriz inversa* de A e escrevemos $B = A^{-1}$.

Muitas das matrizes 2×2 que encontramos no capítulo 1 são inversíveis. Por exemplo, a matriz ρ_θ da rotação anti-horária de um ângulo θ tem $\rho_{2\pi-\theta}$ como inversa porque

$$\rho_\theta \rho_{2\pi-\theta} = \rho_{2\pi-\theta} \rho_\theta = \rho_{2\pi} = I.$$

Já a matriz R de uma reflexão satisfaz

$$R^2 = I,$$

porque se refletimos um vetor duas vezes relativamente a um mesmo espelho obtemos de volta o vetor original. Ambas estas matrizes são exemplos de matrizes ortogonais, que são matrizes cuja inversa coincide com a transposta. Neste artigo veremos como usar eliminação para calcular a inversa de uma matriz que não é ortogonal.

A inversa de uma matriz satisfaz uma propriedade semelhante à transposta. Digamos que A e B são duas matrizes quadradas inversíveis e de mesmo tamanho. Então, pela associatividade do produto de matrizes

$$(AB) \cdot (B^{-1}A^{-1}) = A \cdot (BB^{-1}) \cdot A^{-1} = AA^{-1} = I.$$

Portanto,

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

isto é, a inversa de AB é igual ao produto das inversas de A e B , só que com a ordem dos fatores trocada. Para encerrar provaremos uma propriedade da inversa que será necessária no artigo 2.3.

PROPOSIÇÃO 2.1. *Uma matriz quadrada é inversível se, e somente se, sua transposta for inversível.*

A demonstração consiste, apenas, em observar que se

$$AA^{-1} = I = A^{-1}A$$

então, pelas propriedades da transposta,

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

Portanto, se A é inversível, então A^t é inversível e sua inversa é igual a $(A^{-1})^t$. A recíproca segue do que acabamos de provar e do fato de que $(A^t)^t = A$.

3. Funções de matrizes

3.1. O traço. O *traço* $\text{tr}(A)$ de uma matriz quadrada A é a soma de suas entradas diagonais; isto é

$$\text{tr} \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{bmatrix} = a_{1,1} + \cdots + a_{n,n}.$$

É fácil ver que o traço da soma de duas matrizes é igual à soma dos seus traços. Bem mais útil é o comportamento do traço relativamente à multiplicação de matrizes. Embora o traço do produto não tenha nenhuma relação simples com o produto dos traços, ele satisfaz à seguinte propriedade, que é bastante útil. Se A e B são duas matrizes quadradas $n \times n$, então

$$(56) \quad \text{tr}(AB) = \text{tr}(BA).$$

Para provar isto, digamos que

$$A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{bmatrix} \quad \text{e que} \quad B = \begin{bmatrix} b_{1,1} & \cdots & b_{1,n} \\ \vdots & \ddots & \vdots \\ b_{n,1} & \cdots & b_{n,n} \end{bmatrix}$$

A i -ésima posição da diagonal de AB é obtida multiplicando a i -ésima linha de A pela i -ésima coluna de B , o que nos dá

$$a_{i,1}b_{1,i} + a_{i,2}b_{2,i} + \cdots + a_{i,n}b_{n,i}.$$

Portanto,

$$\text{tr}(AB) = \sum_{i=1}^n \sum_{j=1}^n a_{i,j}b_{j,i}.$$

Analogamente, a j -ésima posição na diagonal de BA é igual a

$$b_{j,1}a_{1,j} + b_{j,2}a_{2,j} + \cdots + b_{j,n}a_{n,j},$$

de modo que

$$\text{tr}(AB) = \sum_{j=1}^n \sum_{i=1}^n b_{j,i}a_{i,j}$$

pode ser obtido do traço de AB simplesmente invertendo a ordem dos somatórios. Há uma consequência importante da fórmula (56) que usaremos mais à frente no livro. Supondo que M e N são matrizes quadradas de mesmo tamanho e que M é inversível, temos que

$$(57) \quad \text{tr}(N) = \text{tr}(MNM^{-1})$$

que é facilmente obtida de (56) tomando-se $A = NM^{-1}$ e $B = M$.

Uma aplicação curiosa do traço está relacionada ao comportamento da posição e do momento linear em mecânica quântica. Na versão inicial da mecânica quântica, publicada em 1925 por W. Heisenberg, o momento P e posição Q de uma partícula são descritos em termos de coeficientes semelhantes aos de uma série de Fourier, mas com a estranha propriedade de que, pelas regras estabelecidas por Heisenberg em seu artigo, $PQ \neq QP$. Na época, Heisenberg era assistente de M. Born em Göttingen e submeteu o artigo a Born, para publicação no *Zeitschrift für Physik*, do qual Born era um dos editores. Como Born relata em sua biografia:

Depois de ter enviado o artigo de Heisenberg para publicação no *Zeitschrift für Physik*, comecei a pensar sobre sua multiplicação simbólica e logo fiquei tão envolvido com ela que pensei sobre ela um dia inteiro e mal consegui dormir à noite. Porque senti que havia algo fundamental por trás dela, que poderia ser a consumação de nosso trabalho de muitos anos. Uma manhã, por volta de 10 de julho de 1925, subitamente a luz se fez: a multiplicação simbólica de Heisenberg nada mais era que o cálculo matricial, que eu conhecia bem desde meus anos de estudante[.]

Na época, ao contrário do que acontece hoje, o cálculo matricial não era parte do conhecimento básico de um físico ou engenheiro; mas a sólida formação matemática de Born ajudou-lhe a enxergar o que Heisenberg não havia visto. Com ajuda de P. Jordan, que também era seu assistente na época, Born deduziu do artigo de Heisenberg que as matrizes P e Q que representam momento e posição deveriam satisfazer

$$(58) \quad PQ - QP = I,$$

em que I é a matriz identidade. A citação acima, assim como mais detalhes sobre os primórdios da mecânica quântica podem ser encontrados em [1, capítulo XIX].

Contudo, pela equação 56,

$$\text{tr}(PQ - QP) = \text{tr}(PQ) - \text{tr}(QP) = 0,$$

ao passo que o traço da matriz identidade é igual ao tamanho n da matriz. A aparente contradição é fácil de explicar: as matrizes que aparecem em mecânica quântica são infinitas. Por exemplo, as matrizes

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 2 & 0 & \dots \\ 0 & 0 & 0 & 3 & \dots \\ 0 & 0 & 0 & 0 & 4 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad \text{e} \quad P = \begin{bmatrix} 0 & 0 & 0 & 0 & \dots \\ 1 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

satisfazem a equação (58) como você pode verificar com um pouco de paciência. O famoso *princípio da incerteza* da Heisenberg é consequência desta equação.

CAPÍTULO 4

Sistemas lineares

Neste capítulo introduzimos um dos mais importantes algoritmos da álgebra linear, usando como motivação sua aplicação à solução de sistemas lineares. Começaremos o capítulo analisando em detalhes um método bem conhecido para a solução de sistemas lineares com apenas duas incógnitas que serve de inspiração para o algoritmo geral.

1. Sistemas com duas variáveis

Nos últimos anos do ensino fundamental aprendemos vários métodos para resolver sistemas lineares de duas variáveis. Um deles, o *método de adição*, consiste em multiplicar uma (ou ambas) as equações por constantes apropriadas de modo que, quando forem somadas, resta uma equação linear em apenas uma das variáveis, que pode então ser facilmente resolvida. Vejamos um exemplo. Se o sistema for

$$\begin{aligned}x + 3y &= 1 \\2x + 5y &= 4,\end{aligned}$$

então subtraímos da segunda equação o dobro da primeira, o que nos dá $-y = 2$; isto é, $y = -2$. Substituindo isto em qualquer das duas equações originais, podemos determinar o valor de x . De fato, da primeira equação

$$x = 1 - 3y = 1 - 3 \cdot (-2) = 7.$$

Portanto o sistema tem solução $x = 7$ e $y = -2$.

Como este método é o ponto de partida para boa parte do que faremos no curso, vamos analisá-lo em detalhe. Começaremos definindo com cuidado algumas noções básicas. Um *sistema linear* nas variáveis x e y corresponde a um par de equações

$$(59) \quad \begin{aligned}a_1x + a'_1y &= b_1 \\a_2x + a'_2y &= b_2;\end{aligned}$$

em que $a_1, a_2, a'_1, a'_2, b_1$ e b_2 são números reais. O mesmo sistema pode ser escrito na forma

$$\begin{aligned}a_1x + a'_1y - b_1 &= 0 \\a_2x + a'_2y - b_2 &= 0;\end{aligned}$$

ou, de maneira ainda mais compacta como

$$(60) \quad E_1 = 0 \quad \text{e} \quad E_2 = 0;$$

em que

$$E_1 = a_1x + a'_1y - b_1 \quad \text{e} \quad E_2 = a_2x + a'_2y - b_2$$

são polinômios lineares. Dados dois números reais x_0 e y_0 , denotaremos por $E_1(x_0, y_0)$ o número real obtido substituindo-se x por x_0 e y por y_0 no polinômio E_1 . No caso do exemplo resolvido no início deste artigo, estes polinômios serão

$$E_1 = x + 3y - 1 \quad \text{e} \quad E_2 = 2x + 5y - 4.$$

Usando esta notação, podemos definir uma *solução do sistema* (60) como sendo um par de números (x_0, y_0) para o qual

$$E_1(x_0, y_0) = E_2(x_0, y_0) = 0.$$

Levando em conta que a ordem das equações não altera o sistema, escolheremos sempre a primeira equação de maneira que nela o x apareça com coeficiente diferente de zero. Observe que esta escolha é sempre possível, porque estamos supondo que se trata de um sistema em *duas* incógnitas. No sistema (59) isto significa que podemos supor que $a_1 \neq 0$ na equação E_1 . A estratégia que adotaremos consiste em substituir o sistema

$$\begin{cases} E_1 = 0 \\ E_2 = 0 \end{cases} \quad \text{por um sistema da forma} \quad \begin{cases} E_1 = 0 \\ cE_1 + E_2 = 0, \end{cases}$$

em que c é um número real. Naturalmente c será escolhido de maneira que o segundo sistema seja mais fácil de resolver que o primeiro. De fato, se $c = -a_2/a_1$, temos que

$$(61) \quad cE_1 + E_2 = \left(a'_2 - \frac{a_2a'_1}{a_1} \right) y - \left(b_2 - \frac{b_1a_2}{a_1} \right);$$

de modo que $cE_1 + E_2 = 0$ é uma equação linear em uma única variável (neste caso y). Se

$$\alpha = a'_2 - \frac{a_2a'_1}{a_1} \quad \text{e} \quad \beta = b_2 - \frac{b_1a_2}{a_1},$$

então

$$(62) \quad 0 = cE_1 + E_2 = \alpha y - \beta.$$

No caso em que $\alpha \neq 0$,

$$y = \frac{\beta}{\alpha}.$$

Para achar o valor de x correspondente, substituímos y por β/α em E_1 , obtendo

$$a_1x + a_2\frac{\beta}{\alpha} = b_1.$$

Note que, como estamos supondo que $a_1 \neq 0$, esta última equação sempre tem uma única solução. Como veremos, este pequeno detalhe é de grande importância para a solução do sistema. Neste caso, resolvendo a equação linear em x obtemos

$$x = -a_2 \frac{\beta}{a_1\alpha} + \frac{b_1}{a_1}.$$

Não podemos esquecer que, embora sempre possamos escolher $a_1 \neq 0$, o mesmo não ocorre com α , que pode muito bem ser nulo. Caso isto aconteça, (62) reduz-se a $0 = \beta$. Portanto, quando $\alpha = 0$, a equação (62) só terá solução se β também for zero. Naturalmente, se $\alpha = \beta = 0$ então qualquer número real serve de solução à equação $\alpha y = \beta$, pois o produto de zero por qualquer escalar é o próprio zero. Mas, como fizemos questão de salientar acima, a cada valor de y corresponde um (único) valor de x , já que $a_1 \neq 0$. Portanto, se (62) tiver infinitas soluções isto implica que o mesmo ocorre com o sistema

$$\begin{aligned} E_1 &= 0 \\ cE_1 + E_2 &= 0, \end{aligned}$$

Esta análise do *método de adição* nos permite formulá-lo como consistindo das seguintes etapas:

Primeira etapa: dado o sistema $E_1 = E_2 = 0$, ordenamos as equações de maneira que o coeficiente de x em E_1 não seja nulo;

Segunda etapa: escolhemos c de modo que $cE_1 + E_2 = 0$ seja uma equação apenas na variável y ;

Terceira etapa: resolvendo a equação linear em uma variável $cE_1 + E_2 = 0$ obtemos os possíveis valores de y ;

Quarta etapa: substituindo cada valor de y obtido na etapa anterior na equação $E_1 = 0$, obtemos uma equação linear, cuja solução nos dá o valor correspondente para x .

A ordenação das equações feita na primeira etapa garante que as soluções do sistema fiquem completamente determinadas pelo resultado da terceira etapa. De fato, como vimos acima, se escrevermos $cE_1 + E_2 = 0$ na forma $\alpha y = \beta$, então

- se $\alpha \neq 0$ o sistema tem uma única solução;
- se $\alpha = \beta = 0$ o sistema tem infinitas soluções;
- se $\alpha = 0$ mas $\beta \neq 0$ o sistema não tem solução.

Antes de encerrar este artigo, há um detalhe muito importante que até agora ignoramos. De fato, ainda que nosso objetivo fosse resolver o sistema $E_1 = E_2 = 0$, o que fizemos foi encontrar as soluções de $E_1 = cE_1 + E_2 = 0$. Naturalmente estes sistemas terem equações

diferentes não é muito significativo, o que importa é que tenham *exatamente as mesmas soluções*, e é isto que provaremos agora. Lembre-se que, segundo a definição dada acima, os números reais x_0 e y_0 definem uma solução do sistema $E_1 = E_2 = 0$ se, e somente se, os números $E_1(x_0, y_0)$ e $E_2(x_0, y_0)$ são ambos nulos. Contudo,

$$E_1(x_0, y_0) = E_2(x_0, y_0) = 0$$

implica que, qualquer que seja $c \in \mathbb{R}$,

$$cE_1(x_0, y_0) + E_2(x_0, y_0) = 0.$$

Como $E_1(x_0, y_0) = 0$ por hipótese, segue-se que x_0 e y_0 também são soluções do sistema $E_1 = cE_1 + E_2 = 0$. Mostramos, assim, que

qualquer solução de $E_1 = E_2 = 0$ também é solução de $E_1 = cE_1 + E_2 = 0$;

falta a recíproca. Para prová-la, suponha que x_1 e y_1 são números reais que definem uma solução de $E_1 = cE_1 + E_2 = 0$. Isto significa que

$$\begin{aligned} E_1(x_1, y_1) &= 0 \\ cE_1(x_1, y_1) + E_2(x_1, y_1) &= 0. \end{aligned}$$

Entretanto, qualquer que seja $c \in \mathbb{R}$,

$$E_2(x_1, y_1) = (cE_1(x_1, y_1) + E_2(x_1, y_1)) - cE_1(x_1, y_1)$$

é igual a zero, já que é uma soma de termos nulos, o que prova a recíproca. Como voltaremos a usar este resultado adiante, vamos enunciá-lo como uma proposição.

PROPOSIÇÃO 1.1. *Sejam E_1 e E_2 polinômios lineares e c um número real. O sistema $E_1 = E_2 = 0$ tem exatamente as mesmas soluções que $E_1 = cE_1 + E_2 = 0$.*

2. Sistemas triangulares

Nesta seção começamos a considerar de que forma o método de adição pode ser generalizado para sistemas maiores, com mais equações e mais incógnitas. Começaremos considerando as etapas três e quatro do método, conforme foi enunciado na página 87, porque precisamos decidir que forma deve ter um sistema linear para que possa ser facilmente resolvido.

Para simplificar a análise suporemos que todos os sistemas sob consideração têm a mesma quantidade de incógnitas e equações. Na verdade, isto não representa uma restrição significativa, porque sempre podemos supor que o sistema tem mais equações, todas nulas, ou mais incógnitas, todas com coeficientes nulos, completando assim o que falta para que o número de incógnitas seja igual ao de equações.

2.1. Sistemas determinados. Lembre-se que, no caso do método de adição, o sistema fácil de resolver a que chegamos tinha:

- (1) uma equação linear nas variáveis x e y ;
- (2) uma equação linear apenas na variável y .

Suponhamos, para simplificar a análise, que a equação linear (2) tenha uma única solução. Para resolver o sistema, determinamos esta solução, que chamaremos de y_0 , a partir de (2) e a substituímos no lugar de y em (1). O resultado é uma equação linear na variável x que, sob as hipóteses feitas na página 87, sempre pode ser resolvida, retornando como solução, um valor x_0 . A solução do sistema será, então, $x = x_0$ e $y = y_0$.

Podemos generalizar isto para um sistema com n equações e n incógnitas x_1, \dots, x_n supondo que

- (63) a k -ésima equação do sistema depende apenas das variáveis x_k, \dots, x_n .

Assim, a primeira equação pode depender de todas as variáveis, a segunda não pode depender da variável x_1 , a terceira não pode depender da variável x_2 , e assim por diante até a n -ésima equação, que dependerá apenas da variável x_n . Um sistema deste tipo é chamado de *triangular superior* por causa da forma que toma quando escrevemos uma equação abaixo da outra. Por exemplo, denotando por $a_{i,j}$ os coeficientes das variáveis e por b_1, \dots, b_n os termos constantes, os sistemas triangulares superiores têm a forma

$$(64) \quad \begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 + a_{1,3}x_3 + \cdots + a_{1,n}x_n &= b_1 \\ a_{2,2}x_2 + a_{2,3}x_3 + \cdots + a_{2,n}x_n &= b_2 \\ a_{3,3}x_3 + \cdots + a_{3,n}x_n &= b_3 \\ &\vdots && \vdots \\ a_{n,n}x_n &= b_n \end{aligned}$$

Sempre que tratarmos de sistemas triangulares superiores, suporemos que suas variáveis foram ordenadas de modo que, na i -ésima equação, os coeficientes de x_1, \dots, x_{i-1} sejam todos nulos, fazendo com que o sistema tenha a forma ilustrada em (64).

Antes de prosseguir, veremos como os sistemas triangulares superiores podem ser resolvidos considerando um exemplo numérico. Seja

$$\begin{aligned} x + 3y + z + w &= 1 \\ 2y - z + 5w &= 6 \\ 5z - w &= 2 \\ 4w &= 12 \end{aligned}$$

Como a quarta e última equação depende apenas de uma variável (neste exemplo w), podemos resolvê-la obtendo $w = 3$. Substituindo este valor para w nas três equações anteriores,

$$\begin{aligned}x + 3y + z + 3 &= 1 \\2y - z + 15 &= 6 \\5z - 3 &= 2,\end{aligned}$$

que pode ser reescrito na forma

$$\begin{aligned}x + 3y + z &= -2 \\2y - z &= -9 \\5z &= 5.\end{aligned}$$

Como este novo sistema linear também é triangular superior, podemos resolver a última equação, obtendo $z = 1$. Substituindo o valor de z nas duas primeiras equações deste sistema,

$$\begin{aligned}x + 3y + 1 &= -2 \\2y - 1 &= -9,\end{aligned}$$

que equivale a

$$\begin{aligned}x + 3y &= -3 \\2y &= -8.\end{aligned}$$

Mais uma vez, trata-se de um sistema triangular superior que, uma vez resolvido, nos dá $y = -4$ e $x = 9$. Portanto, o sistema original tem solução

$$x = 9, \quad y = -4, \quad z = 1 \text{ e } w = 3.$$

Certamente você percebeu que se trata de um procedimento recursivo. Voltando ao caso geral, a solução da última equação nos permite encontrar o valor da variável x_n . Substituindo este valor nas $n - 1$ equações anteriores, o sistema resultante será necessariamente triangular superior. Para provar isto no caso geral, considere novamente o sistema (64). Supondo que $a_{nn} \neq 0$, podemos resolver a última equação, obtendo

$$x_n = \frac{b_n}{a_{n,n}}.$$

Denotando este número por α_n , vamos substituí-lo no lugar de x_n nas $n - 1$ primeiras equações do sistema. Com isto, as parcelas que envolvem x_n nestas equações tornam-se

constantes, ao passo que as parcelas que envolvem x_1, \dots, x_{n-1} não são afetadas. Agrupando os termos constantes do lado direito de cada equação, o sistema resultante é

$$\begin{aligned} A_{1,1}x_1 + A_{1,2}x_2 + A_{1,3}x_3 + \cdots + A_{1,n-1}x_{n-1} &= b_1 - \alpha_n A_{1,n} \\ A_{2,2}x_2 + A_{2,3}x_3 + \cdots + A_{2,n-1}x_{n-1} &= b_2 - \alpha_n A_{2,n} \\ A_{3,3}x_3 + \cdots + A_{3,n-1}x_{n-1} &= b_3 - \alpha_n A_{3,n} \\ &\vdots && \vdots \\ A_{n-1,n-1}x_{n-1} &= b_{n-1} - \alpha_n A_{n-1,n} \end{aligned}$$

Este sistema de $n-1$ equações nas incógnitas x_1, \dots, x_{n-1} é claramente triangular superior. Entretanto, para que a recursão possa de fato ocorrer, é necessário que $a_{i,i} \neq 0$, para todo $1 \leq i \leq n$. Por analogia com as matrizes, diremos que estes são os *coeficientes diagonais* do sistema (64). Por razões óbvias, este algoritmo para a solução de sistemas triangulares superiores é conhecido como *método de substituição*, ou *método de substituição inversa*, já que os valores das incógnitas são substituídos da última para a primeira equação.

O que fizemos até aqui estabelece, entre outras coisas, que o método de substituição nos permite resolver qualquer sistema triangular superior cujos coeficientes diagonais são não nulos. Além disso, o método produz, neste caso, apenas um valor para cada variável, de modo que o sistema é *determinado*.

2.2. Sistemas impossíveis e indeterminados. Vimos no artigo anterior que os sistemas triangulares superiores cujas entradas diagonais não se anulam é determinado, mas o que ocorre se o sistema tiver zeros ao longo da diagonal? Começaremos ilustrando as várias possibilidades com alguns exemplos.

Vejamos o que ocorre se aplicarmos o método de substituição ao sistema

$$(65) \quad \begin{aligned} x + 3y + z + w &= 1 \\ -z + 5w &= 6 \\ 4z - w &= 2 \\ 2w &= 12. \end{aligned}$$

Das duas últimas equações segue facilmente que $w = 6$ e que $z = 2$. Substituindo os valores obtidos para estas duas variáveis nas duas primeiras equações, obtemos

$$\begin{aligned} x + 3y &= -7 \\ 0 &= -22 \end{aligned}$$

que é evidentemente *impossível*. Por outro lado, o sistema

$$\begin{aligned} x + 3y + z + w &= 1 \\ -z + 5w &= 28 \\ 4z - w &= 2 \\ 2w &= 12, \end{aligned}$$

que difere do anterior apenas no lado direito da segunda equação, tem um comportamento totalmente diferente. De fato, substituindo $w = 6$ e $z = 2$ neste segundo sistema, as duas primeiras equações se tornam

$$\begin{aligned} x + 3y &= -7 \\ 0 &= 0. \end{aligned}$$

Em outras palavras, das quatro equações do segundo sistema, só três impõem alguma restrição às variáveis. Com isso podemos achar um valor de x para cada valor que atribuirmos a y ; em outras palavras, o sistema tem infinitas soluções, que podemos descrever parametricamente por

$$x = -7 - 3y, \quad z = 2 \text{ e } w = 6;$$

em que y funciona como um parâmetro; isto é, uma variável que pode tomar qualquer valor que desejemos. Sistemas lineares para os quais existe uma infinidade de soluções são chamados de *indeterminados*. Portanto, a existência de coeficientes nulos ao longo da diagonal faz com que o sistema possa ser impossível ou indeterminado.

Como os exemplos anteriores mostram, a análise do caso em que há coeficientes nulos ao longo da diagonal pode ser um tanto trabalhosa, sobretudo se o sistema tiver muitas equações e incógnitas. Felizmente há uma maneira de contornar este problema, que consiste em exigir que o sistema seja, não apenas triangular superior, mas que satisfaça uma condição ainda mais restritiva.

Dizemos que um sistema está em *forma escada* ou está *escalonado* se o primeiro coeficiente não nulo de cada linha está à direita do primeiro coeficiente não nulo da linha anterior. Mais precisamente, se na k -ésima linha, a variável x_ℓ for a de menor índice cujo coeficiente não é nulo, então as variáveis com coeficientes não nulos na $k+1$ -ésima linha não podem ter índices menores do que $\ell+1$. Apesar de tais sistemas serem claramente triangulares superiores, nem todo sistema triangular superior está na forma escada. Por exemplo,

$$\begin{aligned} x + 3y + z + w &= 1 \\ -z + 5w &= 28 \\ 4z - w &= 2 \\ 2w &= 12, \end{aligned}$$

é triangular superior, mas não está na forma escada. Observe que um sistema na forma escada pode “pular um degrau”, como é o caso de

$$\begin{aligned}x + 3y + z + w &= 1 \\z + 5w &= 28 \\9w &= 2.\end{aligned}$$

Sistemas na forma escada não estão sujeitos a fenômenos como o do exemplo (65), cuja impossibilidade só conseguimos detectar a meio caminho de sua resolução. Isto ocorre porque, um sistema em escada só pode ser impossível se sua última equação não tiver solução. Digamos, por exemplo, que

$$E_1 = 0, \dots, E_n = 0,$$

sejam as equações de um sistema linear e suponhamos que, para um dado inteiro k entre 1 e n , a primeira incógnita com coeficiente não nulo na equação E_k é x_ℓ . Se este sistema está na forma escada, então $\ell \geq k$ e as equações $E_{k+1} = 0, \dots, E_n = 0$ dependem apenas das variáveis

$$x_{\ell+1}, \dots, x_n.$$

Isto significa que, se existirem números reais $\alpha_{\ell+1}, \dots, \alpha_n$ que constituam uma solução de $E_{k+1} = 0, \dots, E_n = 0$, então o valor de x_ℓ pode ser determinado a partir da equação $E_k = 0$, já que x_ℓ tem coeficiente não nulo nesta equação. Como este argumento se aplica a todos os possíveis valores de k para os quais $1 \leq k < n$, somente a última equação pode ser impossível de resolver.

E se o sistema for indeterminado? A única maneira disto acontecer é se a escada que o sistema forma pular um degrau, caso contrário teremos um sistema triangular sem nenhuma posição nula ao longo da diagonal, que será inevitavelmente determinado. Por outro lado, dizer que o sistema pulou um degrau significa que há uma equação a menos intercalada, de modo que o sistema tem menos equações (não nulas) do que incógnitas. As considerações do parágrafo anterior mostram que, neste caso, se a última equação não nula for *possível*, então o sistema terá solução. Por outro lado, como há menos equações que incógnitas, teremos incógnitas sobre as quais não está sendo imposta nenhuma restrição, de modo que o sistema será mesmo indeterminado. Resumimos tudo isto na forma de uma proposição, para referência futura.

PROPOSIÇÃO 2.1. *Seja S um sistema em forma escada com n equações e n incógnitas. Então o sistema é:*

determinado: *se não há posições nulas ao longo da diagonal;*

indeterminado: *se há menos equações não nulas que incógnitas e a última equação diferente de zero tem solução;*

impossível: *se a última equação diferente de zero não tiver solução.*

3. Eliminação gaussiana

Tendo determinado que os sistemas fáceis de resolver são os que estão em forma escada, resta-nos descobrir como proceder para reduzir um sistema geral a esta forma. Começaremos experimentando em alguns exemplos.

3.1. Exemplos. Para nosso primeiro exemplo, consideraremos o sistema linear

$$(66) \quad \begin{aligned} x + y + z + w &= 1 \\ 2x + 3y + z + 5w &= 5 \\ x + 7y - z + 2w &= 3 \\ 5x - y - 3z + w &= 7, \end{aligned}$$

em um sistema em forma escada. Na medida do possível, gostaríamos de usar a mesma estratégia do método de adição: multiplicar uma equação por uma constante cuidadosamente escolhida e somá-la à equação seguinte. Por exemplo, multiplicando a primeira equação do sistema acima por -2 e somando o resultado à segunda equação, obtemos

$$y - z + 3w = 3,$$

que tem uma variável a menos. Substituindo a segunda equação do sistema original por esta, o sistema resultante é

$$\begin{aligned} x + y + z + w &= 1 \\ y - z + 3w &= 3 \\ x + 7y - z + 2w &= 3 \\ 5x - y - 3z + w &= 7. \end{aligned}$$

Para chegar a um sistema triangular superior ainda precisamos eliminar as variáveis x e y na terceira equação e x , y e z na última. Ao invés de tentar fazer isto diretamente, vamos proceder de maneira recursiva. Assim, usando a primeira equação, vamos eliminar o termo em x das duas últimas equações do sistema. Para isto, substituímos a terceira equação por sua soma com o produto da primeira equação por -1 e a quarta equação por sua soma com o produto da primeira equação por -5 , do que resulta o sistema

$$\begin{aligned} x + y + z + w &= 1 \\ y - z + 3w &= 3 \\ 6y - 2z + w &= 2 \\ -6y - 8z - 4w &= 2. \end{aligned}$$

Para transformar este último sistema para a forma triangular superior, basta converter a esta forma o sistema

$$(67) \quad \begin{aligned} y - z + 3w &= 3 \\ 6y - 2z + w &= 2 \\ -6y - 8z - 4w &= 2. \end{aligned}$$

que tem apenas três equações em três incógnitas, uma equação e uma incógnita a menos que o sistema original. Substituindo a segunda equação do sistema (67) por ela própria mais -6 vezes a primeira equação e procedendo de maneira análoga para sua última equação, obtemos

$$\begin{aligned} y - z + 3w &= 3 \\ 4z - 17w &= -16 \\ -14z + 14w &= 20, \end{aligned}$$

e o problema mais uma vez se reduz a transformar em forma triangular um sistema menor, desta vez

$$\begin{aligned} 4z - 17w &= -16 \\ -14z + 14w &= 20, \end{aligned}$$

o que pode ser feito multiplicando por 3 a primeira linha deste sistema dois por $7/2$ e somando o resultado à última linha, do que resulta

$$\begin{aligned} 4z - 17w &= -16 \\ -\frac{91}{2}w &= -36. \end{aligned}$$

Reunindo as equações simplificadas, obtemos o sistema triangular superior

$$(68) \quad \begin{aligned} x + y + z + w &= 1 \\ y - z + 3w &= 3 \\ 4z - 17w &= -16 \\ -\frac{91}{2}w &= -36. \end{aligned}$$

que pode ser facilmente resolvido por substituição, produzindo a solução

$$x = \frac{6}{7}, \quad y = -\frac{1}{91}, \quad z = -\frac{58}{91} \quad \text{e} \quad w = \frac{72}{91}.$$

Uma análise, ainda que superficial, dos cálculos realizados acima mostra que se trata de um procedimento recursivo baseado em uma operação inteiramente similar à utilizada no método de adição. Mais precisamente, dadas duas equações E e E' diremos que a operação que consiste em

escolher uma constante c de modo que $cE + E'$ tenha uma incógnita a menos que E ,

é uma *operação elementar* entre as linhas E e E' . Supondo que o sistema tem n equações em n incógnitas, aplicamos esta operação $n - 1$ vezes, de modo a eliminar uma das incógnitas de $n - 1$ das equações dadas. Estas $n - 1$ equações formam então um sistema com $n - 1$ incógnitas ao qual o mesmo procedimento pode ser aplicado recursivamente. Este algoritmo, conhecido como *método de eliminação de Gauss*, pode ser aplicado a qualquer sistema linear. Como uma incógnita é eliminada a cada passo da recursão, o sistema resultante ao final será triangular superior e, portanto, facilmente resolvido usando os métodos do artigo 2.

Naturalmente, para que os valores retornados por este algoritmo sejam soluções do sistema original é preciso que ele e o sistema triangular superior tenham exatamente as mesmas soluções. Mas isto segue da proposição 1.1, que foi deliberadamente formulada sem que o número de incógnitas nas equações fosse especificado, porque a mesma demonstração funciona se há duas ou qualquer outra quantidade delas presentes nas equações. Talvez você esteja pensando: “tudo bem, mas que só tratamos do caso em que há duas equações, disto não há como escapar”. Sem dúvida, mas na verdade isto basta. Afinal de contas, este algoritmo pode ser considerado como consistindo de uma sucessão de operações elementares aplicadas a vários *pares* de linhas. Naturalmente a palavra chave aqui é *par*: a cada operação realizada, somente uma linha é alterada por outra, as demais ficam como estão. Por isto basta saber que uma solução é comum a duas equações E e E' se, e somente se, é uma solução comum a E e $cE + E'$, que é exatamente o que a proposição 1.1 nos diz.

No próximo artigo veremos uma maneira mais prática de executar este algoritmo. Entretanto, convém desde já chamar a atenção para o fato de que a afirmação, feita acima, de que podemos aplicá-lo a *qualquer sistema linear* deve ser tomada com uma boa dose de cautela. Considere, por exemplo, o sistema

$$\begin{aligned}x + y + z &= 1 \\x + y + 2z &= 3 \\x + 2y - z &= 4.\end{aligned}$$

Usando a primeira equação para eliminar a variável x das outras duas equações, obtemos

$$\begin{aligned}x + y + z &= 1 \\z &= 2 \\y - 2z &= 3.\end{aligned}$$

Note que y não pode ser eliminado da última equação por uma operação elementar, já que seu coeficiente na segunda equação é nulo. Mas por que desejaríamos fazer isto? Afinal, basta trocar as duas últimas equações entre si e o sistema já está na forma triangular superior, o que basta para nossos propósitos. Voltaremos a este ponto no próximo artigo.

3.2. Eliminação em matrizes. Ao contrário do que aconteceu no método de solução de sistemas triangulares superiores por substituição, as variáveis desempenham um papel totalmente secundário no algoritmo de eliminação. De fato, nunca substituímos nenhum valor nas variáveis, elas servem apenas como marcadores de posição, para sabermos quais coeficientes podem ser somados com quais outros coeficientes. Mas, sejamos realistas, para isto não precisamos escrever as variáveis, basta-nos escrever os coeficientes de uma mesma variável sistematicamente uns sobre os outros numa tabela ou, usando a terminologia da álgebra linear, numa matriz, conhecida como *matriz aumentada do sistema*. Por exemplo,

$$(69) \quad \begin{aligned} x + y + z + w &= 1 \\ 2x + 3y + z + 5w &= 5 \\ x + 7y - z + 2w &= 3 \\ 5x - y - 3z + w &= 7, \end{aligned}$$

tem como matriz aumentada

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

Agora, executando uma operação elementar entre as duas primeiras equações de (69), obtemos o sistema

$$\begin{aligned} x + y + z + w &= 1 \\ y - z + 3w &= 3 \\ x + 7y - z + 2w &= 3 \\ 5x - y - 3z + w &= 7, \end{aligned}$$

cuja matriz aumentada é

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

A operação elementar que efetuamos sobre as duas primeiras equações de (69) pode ser facilmente reformulada em termos das linhas da matriz. Considerando as linhas de A como matrizes 1×5 , a operação executada substitui a segunda linha por ela própria, somada a -2 vezes a primeira linha:

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

Em geral, se L e L' são duas linhas de uma matriz M e c é um número real, diremos que

ao substituir L' por $L' + cL$ em M efetuamos uma *operação elementar por linha*.

A não ser que c ou L sejam nulos, a matriz que resulta da aplicação de uma operação elementar por linha a M é diferente de M . Naturalmente, quando aplicamos uma operação elementar por linha como parte do método de eliminação, escolhemos a constante c de modo a eliminar a posição não nula mais à esquerda de L' , *desde que isto seja possível*. A passagem de uma matriz à outra, efetuada a partir de uma operação elementar por linha, será denotada por uma seta \rightarrow . Por exemplo,

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

no caso da eliminação efetuada acima. Dando continuidade à aplicação da versão matricial do processo de eliminação a este exemplo, temos

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 3 & 3 \\ 1 & 7 & -1 & 2 & 3 \\ 5 & -1 & -3 & 1 & 7 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 3 & 3 \\ 0 & 6 & -2 & 1 & 2 \\ 5 & -1 & -3 & 1 & 7 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 3 & 3 \\ 0 & 6 & -2 & 1 & 2 \\ 0 & -6 & -8 & -4 & 2 \end{bmatrix},$$

com o quê todas as posições abaixo de 1, 1 são agora nulas. A propósito, a entrada da posição 1, 1 é conhecida como o *pivô* desta etapa da eliminação. Em geral, o *pivô de uma linha* é a entrada não nula mais à esquerda desta linha. Quando o pivô da i -ésima linha está na posição i, j usamos operações elementares por linha para anular todas as posições da j -ésima coluna de uma matriz que ficam abaixo de sua i -ésima linha; isto é, aquelas localizadas nas linhas $i + 1$ até n .

Passando à segunda coluna, usamos a entrada da posição 2, 2 como pivô para anular, em dois passos, cada uma das posições desta coluna que ficam abaixo de 2, 2, como segue

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 3 & 3 \\ 0 & 6 & -2 & 1 & 2 \\ 0 & -6 & -8 & -4 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 3 & 3 \\ 0 & 0 & 4 & -17 & -16 \\ 0 & -6 & -8 & -4 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 3 & 3 \\ 0 & 0 & 4 & -17 & -16 \\ 0 & 0 & -14 & 14 & 20 \end{bmatrix}$$

Finalmente a entrada da posição 3, 3 é usada como pivô para eliminar 4, 3, e resta a matriz

$$\left[\begin{array}{ccccc} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 3 & 3 \\ 0 & 0 & 4 & -17 & -16 \\ 0 & 0 & 0 & -91/2 & -36 \end{array} \right]$$

da qual o sistema triangular superior pode ser facilmente obtido, bastando para isto lembrar que os coeficientes das variáveis foram escritos nas colunas na ordem em que as variáveis aparecem no sistema. Portanto, a primeira coluna corresponde aos coeficientes de x , a segunda aos de y , e assim por diante. Fazendo isto, obtemos o sistema (68) da página 95, como você pode facilmente verificar.

É importante entender que não há nada a *provar* sobre a versão matricial do método de eliminação. Afinal, do ponto de vista matemático, a única coisa que fizemos foi escrever os coeficientes do sistema em uma matriz, em vez de usar as variáveis como marcadores de posição, como vínhamos fazendo até aqui. A operação elementar por linha não passa de uma transcrição direta para a linguagem das matrizes da operação elementar sobre as equações introduzida no artigo 3.1. Porém, uma vez que o método foi traduzido em termos matriciais, nada nos impede de aplicá-lo a qualquer matriz, mesmo uma que não seja matriz aumentada de nenhum sistema.

Para um outro exemplo, um pouco menos ingênuo, considere o sistema

$$\begin{aligned} x + 4y + 6z + w &= 11 \\ 2x + 8y + 5z - w &= 9 \\ 3x + 5y + 2z - 5w &= 5 \\ 4x + 2y - z - 3w &= 1 \end{aligned}$$

cuja matriz aumentada é

$$A = \left[\begin{array}{cccc|c} 1 & 4 & 6 & 1 & 11 \\ 2 & 8 & 5 & -1 & 9 \\ 3 & 5 & 2 & -5 & 5 \\ 4 & 2 & -1 & -3 & 1 \end{array} \right].$$

Desta vez, usamos uma barra vertical para separar os coeficientes das variáveis dos termos constantes. Esta barra não é parte da matriz, mas sim um recurso visual que nos ajuda a distinguir dois grupos de números com significados distintos e evitar confusões e erros. Um computador, obviamente, não requer tais artifícios.

Usando operações elementares por linha para anular as posições da primeira coluna de A abaixo de 1, 1, obtemos a matriz

$$\left[\begin{array}{cccc|c} 1 & 4 & 6 & 1 & 11 \\ 0 & 0 & -7 & -3 & -13 \\ 0 & -7 & -16 & -8 & -28 \\ 0 & -14 & -25 & -7 & -43 \end{array} \right].$$

Estritamente falando, o processo de eliminação descrito anteriormente não pode continuar a partir deste ponto. De fato, isto requereria usar como pivô a entrada da posição 2, 2 que, na matriz acima, é igual a zero. Entretanto, estas linhas são apenas uma representação abreviada das equações do sistema original. Como trocar a ordem das equações não afeta o conjunto solução de um sistema, estamos livres para permutar as linhas de qualquer maneira que desejarmos. No caso da matriz acima, reposicionaremos a segunda linha, que será movida para o lugar da última, de modo que a matriz se torna

$$\left[\begin{array}{cccc|c} 1 & 4 & 6 & 1 & 11 \\ 0 & -7 & -16 & -8 & -28 \\ 0 & -14 & -25 & -7 & -43 \\ 0 & 0 & -7 & -3 & -13 \end{array} \right],$$

com o quê o processo de eliminação pode continuar. Tomando a entrada -7 da posição 2, 2 como pivô, anulamos a posição imediatamente abaixo dela, obtendo

$$\left[\begin{array}{cccc|c} 1 & 4 & 6 & 1 & 11 \\ 0 & -7 & -16 & -8 & -28 \\ 0 & 0 & 7 & 9 & 13 \\ 0 & 0 & -7 & -3 & -13 \end{array} \right].$$

Finalmente, usando a entrada 7 da posição 3, 3 como pivô, obtemos

$$\left[\begin{array}{cccc|c} 1 & 4 & 6 & 1 & 11 \\ 0 & -7 & -16 & -8 & -28 \\ 0 & 0 & 7 & 9 & 13 \\ 0 & 0 & 0 & 6 & 0 \end{array} \right],$$

que é triangular superior. O sistema triangular correspondente, cujas equações são,

$$\begin{aligned}x + 4y + 6z + w &= 11 \\ -7y - 16z - 8w &= -28 \\ 7z + 9w &= 13 \\ 6w &= 0\end{aligned}$$

tem solução

$$x = \frac{41}{49}, \quad y = -\frac{12}{49}, \quad z = \frac{13}{7}, \quad w = 0.$$

Ainda há um detalhe importante sobre a eliminação que não apareceu nos exemplos anteriores. Considere o sistema

$$\begin{aligned}x + 3y + z + w &= 1 \\ x + 3y + 6w &= 29 \\ x + 3y + 4z + 5w &= 31 \\ x + 3y + 16z + 2w &= 37\end{aligned}$$

cuja matriz aumentada é

$$\left[\begin{array}{cccc|c} 1 & 3 & 1 & 1 & 1 \\ 1 & 3 & 0 & 6 & 29 \\ 1 & 3 & 4 & 5 & 31 \\ 1 & 3 & 16 & 2 & 37 \end{array} \right]$$

aplicando eliminação gaussiana a esta matriz, obtemos

$$\left[\begin{array}{cccc|c} 1 & 3 & 1 & 1 & 1 \\ 0 & 0 & -1 & 5 & 28 \\ 0 & 0 & 3 & 4 & 30 \\ 0 & 0 & 15 & 1 & 36 \end{array} \right] \rightarrow \left[\begin{array}{cccc|c} 1 & 3 & 1 & 1 & 1 \\ 0 & 0 & -1 & 5 & 28 \\ 0 & 0 & 3 & 4 & 30 \\ 0 & 0 & 0 & 6 & 42 \end{array} \right] \rightarrow \left[\begin{array}{cccc|c} 1 & 3 & 1 & 1 & 1 \\ 0 & 0 & -1 & 5 & 28 \\ 0 & 0 & 0 & 19 & 114 \\ 0 & 0 & 0 & 76 & 456 \end{array} \right]$$

Como a última linha é igual a quatro vezes a anterior, obtemos, finalmente

$$\left[\begin{array}{cccc|c} 1 & 3 & 1 & 1 & 1 \\ 0 & 0 & -1 & 5 & 28 \\ 0 & 0 & 0 & 19 & 114 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

que corresponde ao sistema triangular superior

$$\begin{aligned}x + 3y + z + w &= 1 \\ -z + 5w &= 28 \\ 19w &= 114,\end{aligned}$$

cujas soluções são

$$x = -7 - 3y, \quad z = 2, \quad w = 6;$$

de modo que o sistema é indeterminado.

Observe que se trata, na verdade, de um sistema em forma escada, o que sugere a seguinte definição. Uma matriz A de tamanho $n \times n$ está em *forma escada*, ou é *escalonada por linhas*, quando a seguinte condição for satisfeita para todo $1 \leq i \leq n$:

se a primeira entrada não nula da i -ésima linha está na j -ésima coluna, então a da $i + 1$ -ésima linha não pode aparecer antes da $j + 1$ -ésima coluna.

Naturalmente, a *primeira entrada não nula* de uma dada linha é aquela entrada não nula que aparece mais à esquerda naquela linha. Aplicada a uma matriz qualquer, a eliminação gaussiana sempre retorna uma matriz em forma escada. De fato, se as linhas i e j de uma matriz A têm primeira entrada não nula na coluna k , podemos usar $A_{i,k}$ como pivô para anular $A_{j,k}$, encurtando assim uma das linhas.

3.3. Outros exemplos. A melhor maneira de descrever exemplos variados de sistemas lineares é tentar entender como um sistema se comporta quando alteramos os valores dos seus coeficientes. Para tornar isto viável do ponto de vista prático, permitiremos que variem apenas alguns coeficientes, que são conhecidos como os *parâmetros* do sistema.

Por exemplo, dado que o sistema

$$\begin{cases} x - ky + z &= 0 \\ kx + (1 - k^2)y + (1 + k)z &= k \\ kx - ky + z &= 1 - k \end{cases}$$

tem k como seu único parâmetro, podemos nos perguntar para que valores de k o sistema é determinado, indeterminado ou impossível. A matriz aumentada deste sistema, que é

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

tem forma escalonada igual a

$$\left[\begin{array}{ccc|cc} 1 & -k & 1 & 0 & \\ 0 & 1 & 1 & k & \\ 0 & 0 & 1-k^2 & -k^3+k^2-k+1 & \end{array} \right],$$

que corresponde ao sistema triangular superior

$$\begin{aligned} x - ky + z &= 0 \\ y + z &= k \\ (1 - k^2)z &= -k^3 + k^2 - k + 1. \end{aligned}$$

Note que se $k = 1$ todos os coeficientes da última equação se anulam, de modo que, neste caso, o sistema terá infinitas soluções; uma para cada valor que escolhermos para z . Por outro lado, se $k = -1$, então a última equação se torna $0 = 4$, o que torna o sistema impossível neste caso. Finalmente, se $k \neq \pm 1$ a última equação tem uma única solução, a saber

$$z = \frac{-k^3 + k^2 - k + 1}{(1 - k^2)};$$

da qual podemos deduzir valores para y e x usando o método de substituição. Resumindo, o sistema é:

- determinado se $k \neq \pm 1$;
- indeterminado se $k = 1$;
- impossível se $k = -1$.

Note que, neste exemplo, o comportamento deste sistema ficou completamente determinado por sua última equação. Isto porque, tendo escolhido k de modo que exista um valor de z que seja solução da última equação, valores correspondentes para x e y sempre podem ser encontrados. Mais precisamente, embora os valores de x e y possam depender de k ; a possibilidade de encontrá-los não é afetada pelo valor de k escolhido. Como mostra nosso próximo exemplo, nem sempre as coisas são tão simples.

Mais uma vez trata-se de um sistema a um parâmetro, cujas equações são

$$\begin{aligned} x + ky + 7z + 9w &= k \\ 3x + (4k + 1)y + 22z + 28w &= 3k + 3 \\ 2x + (3k + 1)y + (2k + 15)z + 20w &= 2k - 5 \\ x + (3k + 2)y + (2k + 9)z + (k + 10)w &= k + 30. \end{aligned}$$

Aplicando eliminação gaussiana à matriz aumentada

$$\left[\begin{array}{cccc|c} 1 & k & 7 & 9 & k \\ 3 & (4k+1) & 22 & 28 & 3k+3 \\ 2 & (3k+1) & (2k+15) & 20 & 2k-5 \\ 1 & (3k+2) & (2k+9) & (k+10) & k+30 \end{array} \right]$$

obtemos a matriz escalonada

$$\left[\begin{array}{cccc|c} 1 & k & 7 & 9 & k \\ 0 & k+1 & 1 & 1 & 3 \\ 0 & 0 & 2k & 1 & -8 \\ 0 & 0 & 0 & k-2 & 16 \end{array} \right]$$

que corresponde ao sistema triangular superior

$$\begin{aligned} x + ky + 7z + 9w &= k \\ (k+1)y + z + w &= 3 \\ 2kz + w &= -8 \\ (k-2)w &= 16 \end{aligned}$$

Para que a última equação tenha solução é preciso que $k \neq 2$. Portanto, se $k = 2$ já temos que o sistema é impossível. Entretanto, como há k em outras posições da diagonal, a análise precisa continuar. Se $k = 0$, o sistema se torna

$$\begin{aligned} x + 7z + 9w &= 0 \\ y + z + w &= 3 \\ w &= -8 \\ -2w &= 16; \end{aligned}$$

que é indeterminado, pois as duas últimas equações coincidem e nenhuma restrição é imposta a z . Finalmente, se $k = -1$, o sistema é

$$\begin{aligned} x - y + 7z + 9w &= k \\ z + w &= 3 \\ -2z + w &= -8 \\ -3w &= 16 \end{aligned}$$

que nos dá $w = 16/3$. Como

$$z = \frac{1}{2} \left(8 + \frac{16}{3} \right) = \frac{20}{3}$$

da penúltima equação, ao passo que

$$z = 3 - \frac{16}{3} = \frac{7}{3}$$

da antepenúltima equação, podemos concluir que o sistema é impossível neste caso. Portanto, o sistema é

determinado: se $k \neq -1, 0, 2$;

impossível: se $k = 2$ ou $k = -1$;

indeterminado: se $k = 0$.

Finalmente, considere o sistema a três parâmetros

$$\begin{cases} x + 8y - 2z = a \\ 5x + 4y - 2z = b \\ 7x - 16y + 2z = c. \end{cases}$$

Desta vez, queremos c em função de a e b de modo que o sistema seja determinado, indeterminado ou impossível. A matriz aumentada

$$\left[\begin{array}{ccc|c} 1 & 8 & -2 & a \\ 5 & 4 & -2 & b \\ 7 & -16 & 2 & c \end{array} \right]$$

tem forma escalonada igual a

$$\left[\begin{array}{ccc|c} 1 & 8 & -2 & a \\ 0 & -36 & 8 & b - 5a \\ 0 & 0 & 0 & 3a - 2b + c \end{array} \right]$$

à qual corresponde o sistema triangular superior

$$\begin{aligned} x + 8y - 2z &= a \\ -36y + 8z &= b - 5a \\ 0 &= 3a - 2b + c \end{aligned}$$

Para começar, o sistema nunca tem uma única solução, porque a única maneira da última equação fazer sentido é se seu termo constante for igual a zero. Isto ocorre quando $c = -3a + 2b$, de modo que, sob esta condição o sistema tem solução; na verdade, infinitas soluções. Resumindo, o sistema é:

determinado: nunca;

impossível: quando $c \neq -3a + 2b$;

indeterminado: quando $c = -3a + 2b$.

4. Aplicações

4.1. Interpolação polinomial. Seja P um conjunto *finito* de pontos do plano. Em uma primeira aproximação, o problema da *interpolação polinomial* almeja obter uma função polinomial

$$f(x) = a_n x^n + \cdots + a_0,$$

de grau n e coeficientes reais, cujo gráfico contém todos os pontos de P . À primeira vista a restrição a funções polinomiais pode parecer muito forte. Entretanto, estas são as únicas funções que um computador consegue calcular diretamente, porque são definidas por uma quantidade finita de operações aritméticas elementares. Para as demais funções, teremos que utilizar aproximações que, frequentemente, são aproximações polinomiais.

Comecemos com um exemplo simples. Digamos que queremos uma função polinomial de grau dois que passa pelos pontos $(1, 0)$, $(2, 1)$ e $(3, 4)$ do plano. A função pode ser escrita na forma

$$f(x) = ax^2 + bx + c,$$

em que a , b e c denotam números reais. Nossa objetivo é determinar quais devem ser os coeficientes a , b e c para que $y = f(x)$ passe pelos pontos dados. Mas para que isto ocorra, devemos ter que

$$\begin{aligned} a + b + c &= f(1) = 0 \\ 4a + 2b + c &= f(2) = 1 \\ 9a + 3b + c &= f(3) = 4 \end{aligned}$$

que podemos interpretar como um sistema linear, cuja matriz aumentada é

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 4 & 2 & 1 & 1 \\ 9 & 3 & 1 & 4 \end{array} \right].$$

Aplicando eliminação gaussiana a esta matriz, obtemos a matriz escada

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

que corresponde ao sistema triangular superior

$$\begin{aligned} a + b + c &= 0 \\ -2b - 3c &= 1 \\ c &= 1 \end{aligned}$$

cuja solução é $c = 1$, $b = -2$ e $a = 1$. Portanto, o polinômio desejado é

$$f(x) = x^2 - 2x + 1,$$

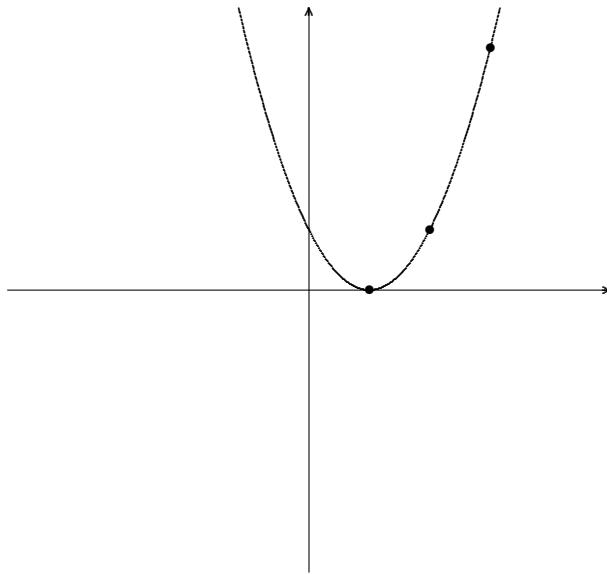


FIGURA 1. Parábola

cujo gráfico, esboçado na figura 1, realmente passa pelos pontos dados.

Esta maneira de executar a interpolação é conhecida como *método dos coeficientes a determinar*, por razões óbvias. O algoritmo geral para este método pode ser descrito da seguinte maneira.

ALGORITMO 4.1. *Seja P um conjunto finito de pontos do plano e n um número positivo, o algoritmo retorna um polinômio de grau n cujo gráfico passa por todos os pontos de P ou uma mensagem de erro.*

- Initialize $S = \emptyset$;
- seja

$$f(x) = a_n x^n + \cdots + a_1 x + a_0$$

um polinômio de grau n cujos coeficientes são valores a determinar (variáveis);

- Para cada ponto $(x_0, y_0) \in P$ construa a equação $f(x_0) = y_0$ e acrescente-a ao sistema S ;
- resolva o sistema S ;
- se o sistema for determinado ou indeterminado, atribua os valores de uma solução aos coeficientes de f e retorne o resultado;
- se o sistema for impossível, retorne uma mensagem de erro.

Problemas de interpolação surgem mesmo no caso em que a curva a ser determinada não é uma função. Por exemplo, dados os mesmos três pontos do exemplo acima, podemos

determinar um círculo que passe por todos os três. Tal círculo terá equação

$$(x - a)^2 + (y - b)^2 = r^2$$

que, uma vez expandida, corresponde a

$$x^2 - 2ax + a^2 + y^2 - 2by + b^2 = r^2;$$

ou ainda a

$$x^2 - 2ax + y^2 - 2by = r^2 - a^2 - b^2.$$

Escrevendo $c = r^2 - a^2 - b^2$, determinaremos os valores de a , b e c para os quais a curva

$$x^2 - 2ax + y^2 - 2by - c = 0$$

contém os pontos dados, no nosso exemplo $(1, 0)$, $(2, 1)$ e $(3, 4)$. Fazendo as devidas substituições, obtemos o sistema

$$\begin{aligned} 2a + c &= 1 \\ 4a + 2b + c &= 5 \\ 6a + 8b + c &= 25; \end{aligned}$$

cuja solução é $a = -2$, $b = 4$ e $c = 5$. Como

$$r^2 = c + a^2 + b^2 = 5 + 16 + 4 = 25$$

a solução do problema é uma circunferência de raio 5 com centro no ponto $(-2, 4)$. Mais uma vez, convém observar que se estivéssemos tentando achar uma circunferência que passasse por quatro, em vez de três, pontos dados, provavelmente teríamos um sistema impossível. Afinal, como aprendemos em geometria elementar, três pontos não alinhados bastam para determinar uma circunferência.

Exercícios

1. Resolva cada um dos sistemas abaixo pelo método de adição.

$$\begin{array}{lll} \text{(a)} \quad \begin{cases} x - 2y = -24 \\ 2x + 3y = 2 \end{cases} & \text{(b)} \quad \begin{cases} 7x + 6z = 1 \\ 2x + 3y = -2 \end{cases} & \text{(c)} \quad \begin{cases} x - 7y = 12 \\ 4x + 16y = 16 \end{cases} \\ \text{(d)} \quad \begin{cases} 3x + y = 0 \\ 9x + 3y = 0 \end{cases} & \text{(e)} \quad \begin{cases} x - y - 2z = 0 \\ 3x - 3y = 20 \end{cases} & \text{(f)} \quad \begin{cases} -3x + 3y = -2 \\ 5x + 2y = 1 \end{cases} \end{array}$$

2. Resolva cada um dos sistemas triangulares abaixo pelo método substituição.

$$(a) \begin{cases} x - 2y - 7z = -24 \\ 3y - 2z = 2 \\ 4z = 5 \end{cases} \quad (b) \begin{cases} x + 4y + 6z = 11 \\ 9y + 7z = 9 \\ z = 7 \end{cases}$$

$$(c) \begin{cases} x - y - 2z - w = 0 \\ 5y + 3z + w = 0 \\ -z - w = 0 \end{cases} \quad (d) \begin{cases} x + 2y - w = 0 \\ y + 2z - w = 0 \\ 2z - w = 0 \\ -3w = 6 \end{cases}$$

3. Resolva cada um dos sistemas abaixo pelo método de eliminação gaussiana.

$$(a) \begin{cases} x - 2y - 7z = -24 \\ 2x + 3y - 2z = 2 \\ 3x - 5y + 4z = 5 \end{cases} \quad (b) \begin{cases} x + 4y + 6z = 11 \\ 2x + 3y + 4z = 9 \\ 3x + 2y + 2z = 7 \end{cases}$$

$$(c) \begin{cases} x + 7y + 14z = 20 \\ 3x + 9y + 12z = 24 \\ 4x + 16y + 26z = 46 \end{cases} \quad (d) \begin{cases} 3x + y + 2z = 0 \\ 9x + 3y - z = 0 \\ 3x + 2y - 3z = 0 \end{cases}$$

$$(e) \begin{cases} x - y - 2z - w = 0 \\ 3x + y + 3z + w = 0 \\ x - y - z - 5w = 0 \end{cases} \quad (f) \begin{cases} -3x + 3y + 2z + w = -2 \\ 5x + 2y + z - 2w = 1 \\ 2x + 5y + 3z - w = -1 \end{cases}$$

$$(g) \begin{cases} x + 2y - w = 0 \\ x + 2z - w = 0 \\ x + 2y + 2z - w = 0 \\ 3x + 4y + 4z - 3w = 0 \end{cases}$$

4. Resolva cada um dos sistemas abaixo pelo método de eliminação, indicando se são determinados, indeterminados ou impossíveis.

$$(a) \begin{cases} x - y - z + w = 1 \\ 2x + 2y - 3z + 6w - u = -1 \\ x - 2y - z + 2w - u = 0 \\ 3x + y - 4z + 7w - u = 0 \end{cases} \quad (b) \begin{cases} x - 3y - 2z + w = 1 \\ x + 2y - 3z + 6w - u = 1 \\ 5x - 2y - z + 2w - 8u = 0 \\ 3x + 2y - 4z + 7w - u = 0 \end{cases}$$

5. Suponhamos que estamos para aplicar eliminação gaussiana a partir da k -ésima linha de uma matriz A de tamanho $n \times n$:

(a) mostre que o pivô tem que estar na linha k, ℓ em que $\ell \geq k$;

- (b) construa um exemplo em que $\ell > k$.
6. Mostre que a quantidade máxima de operações por linha necessárias para transformar uma matriz $n \times n$ dada em sua forma escada por eliminação gaussiana é igual a $n(n - 1)/2$.
7. Determine os valores de k para os quais os sistemas abaixo são determinados, indeterminados ou impossíveis.

$$(a) \begin{cases} x + y + z = 0 \\ x - y + kz = 2 \\ kx + 2y + z = 0 \end{cases}$$

$$(b) \begin{cases} x + y = k \\ k^2x + y = k \end{cases}$$

$$(c) \begin{cases} x + y + kz = 2 \\ 3x + 4y + 2z = k \\ 2x + 3y + z = 1 \end{cases}$$

$$(d) \begin{cases} x + 2y + 3z = 1 \\ x + (2k+4)y + 5z = 5 \\ 2x + (3k+7)y + (k+7)z = k^2 + 4 \end{cases}$$

8. Determine uma função polinomial que passa pelos pontos $(1, 1)$, $(2, 3)$, $(3, 5)$ e $(4, 4)$ para cada um dos seguintes graus: 2, 3 e 4.
9. Determine uma circunferência que passa por cada um dos seguintes conjuntos de pontos do plano:
- (a) $(1, 1)$, $(2, 3)$, $(3, 5)$ e $(4 + \sqrt{2}, 4)$;
- (b) $(2, 5)$, $(3, 4)$, $(4, 5)$ e $(5/2, (10 + \sqrt{3})/2)$.

CAPÍTULO 5

Determinante e matrizes inversas

Apesar de ter-se tornado impopular em anos recentes, o determinante é uma ferramenta extremamente útil, com aplicações em áreas muito além da álgebra linear. A impopularidade dos determinantes nas décadas de 1950 a 1970 deveu-se em grande parte ao forte viés abstrato da matemática naquelas décadas. Com a popularização dos computadores a partir dos anos 1980, seria de esperar que a importância dos determinantes voltasse a ser reconhecida, o que de fato aconteceu no âmbito da computação simbólica. Infelizmente, muitos livros recentes de álgebra linear foram escritos do ponto de vista da análise numérica, área em que a utilidade dos determinantes é bem mais limitada, gerando, mais uma vez, uma falsa impressão quanto à sua utilidade.

1. Determinantes

Há muitas maneiras de definir determinantes, todas equivalentes. A que vamos utilizar é recursiva: ela reduz o cálculo do determinante de uma matriz $n \times n$ ao cálculo do determinante de várias matrizes $(n - 1) \times (n - 1)$.

1.1. Definição e exemplos. O passo indutivo utiliza os determinantes dos cofatores da matriz. Seja A uma matriz $n \times n$ e seja $1 \leq j \leq n$ um número inteiro. O j -ésimo cofator de A é a matriz \hat{A}_j de tamanho $(n - 1) \times (n - 1)$ obtida removendo-se de A sua primeira linha e j -ésima coluna. Por exemplo, se

$$(70) \quad A = \begin{bmatrix} 1 & 2 & 3 \\ 6 & 8 & 9 \\ 1 & 0 & 5 \end{bmatrix}$$

então,

$$\hat{A}_1 = \begin{bmatrix} 8 & 9 \\ 0 & 5 \end{bmatrix}, \quad \hat{A}_2 = \begin{bmatrix} 6 & 9 \\ 1 & 5 \end{bmatrix} \quad \text{e} \quad \hat{A}_3 = \begin{bmatrix} 6 & 8 \\ 1 & 0 \end{bmatrix}$$

são os três cofatores de A . Denotando as entradas de A por $a_{i,j}$, para $1 \leq i \leq n$ e $1 \leq j \leq n$, o *determinante* $\det(A)$ é definido recursivamente por

$$\det(A) = a_{1,1}$$

se $n = 1$ e

$$(71) \quad \det(A) = a_{1,1} \det(\widehat{A}_1) - a_{1,2} \det(\widehat{A}_2) + a_{1,3} \det(\widehat{A}_3) - \cdots + (-1)^{n-1} a_{1,n} \det(\widehat{A}_n),$$

se $n \geq 2$. Note que o sinal das parcelas se alternam na soma acima. Podemos reescrever a fórmula indutiva (71) de forma mais compacta usando somatórios

$$(72) \quad \det(A) = \sum_{j=1}^n (-1)^{j-1} a_{1,j} \det(\widehat{A}_j).$$

Começaremos mostrando que a definição recursiva produz a fórmula usual para o determinante de uma matriz 2×2 . Os cofatores de

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

são

$$\widehat{A}_1 = [a_{2,2}] \quad \text{e} \quad \widehat{A}_2 = [a_{2,1}],$$

de modo que a fórmula recursiva nos dá

$$\det(A) = a_{1,1} \det(\widehat{A}_1) - a_{1,2} \det(\widehat{A}_2) = a_{1,1}a_{2,2} - a_{1,2}a_{2,1},$$

pois todos os dois cofatores são matrizes 1×1 .

Para nosso segundo exemplo, calcularemos o determinante da matriz A em (70). Como já sabemos que a fórmula recursiva coincide com a fórmula usual quando a matriz é 2×2 , usaremos esta fórmula diretamente para simplificar uma parte dos cálculos. Fazendo isso, os determinantes dos cofatores de A são,

$$\det(\widehat{A}_1) = 40, \quad \det(\widehat{A}_2) = 21 \quad \text{e} \quad \det(\widehat{A}_3) = -8.$$

Então, pela fórmula recursiva,

$$\det(A) = 1 \cdot \det(\widehat{A}_1) - 2 \cdot \det(\widehat{A}_2) + 3 \cdot \det(\widehat{A}_3),$$

onde

$$\det(A) = 1 \cdot 40 - 2 \cdot 21 + 3 \cdot (-8) = -26.$$

Nos próximos artigos investigaremos várias propriedades do determinante, mas um caso especial de uma delas é simples o suficiente para que possamos enunciá-lo e prová-lo ainda neste artigo.

LEMA 1.1. *Se a primeira linha de uma matriz é multiplicada por uma constante, então seu determinante é multiplicado pela mesma constante.*

Para provar isto basta observar que cada parcela da soma na equação (71) é multiplicada por exatamente uma entrada da primeira linha da matriz A .

1.2. Troca de linhas. Ao longo deste artigo suporemos que

$$(73) \quad A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ a_{3,1} & a_{3,2} & \cdots & a_{3,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{bmatrix},$$

é uma matriz $n \times n$.

PROPRIEDADE 1.2. *O determinante da matriz obtida trocando-se duas linhas de A é igual a $-\det(A)$.*

Dado que a definição de determinante é recursiva, é natural provar esta propriedade usando indução. Se $n = 1$ não há nada a fazer. Se $n = 2$, então, como vimos no artigo anterior,

$$\det \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} = a_{1,1}a_{2,2} - a_{1,2}a_{2,1}.$$

Mas, trocando as linhas de posição, obtemos

$$\det \begin{bmatrix} a_{2,1} & a_{2,2} \\ a_{1,1} & a_{1,2} \end{bmatrix} = a_{1,2}a_{2,1} - a_{1,1}a_{2,2} = -\det \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix},$$

provando, assim, a base da indução. Suponhamos, então, que $n \geq 3$ e que o resultado vale para todas as matrizes $(n-1) \times (n-1)$. Vamos analisar o que ocorre quando trocamos a k -ésima linha de A com sua ℓ -ésima linha. Neste ponto a demonstração se subdivide em três casos.

O primeiro caso é aquele em que $k, \ell \neq 1$. Neste caso, o j -ésimo cofator de A será

$$\hat{A}_j = \begin{bmatrix} a_{2,1} & \cdots & a_{2,j-1} & a_{2,j+1} & \cdots & a_{2,n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{k,1} & \cdots & a_{k,j-1} & a_{k,j+1} & \cdots & a_{k,n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{\ell,2} & \cdots & a_{\ell,j-1} & a_{\ell,j+1} & \cdots & a_{\ell,n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,j-1} & a_{n,j+1} & \cdots & a_{n,n} \end{bmatrix}$$

Portanto, se B for a matriz obtida trocando-se as linhas k e ℓ de A , seu j -ésimo cofator \hat{B}_j será igual a \hat{A}_j com suas linhas $k-1$ e $\ell-1$ trocadas entre si. Logo, pela hipótese de indução,

$$\det(\hat{B}_j) = -\det(\hat{A}_j),$$

para todo $1 \leq j \leq n$. Temos, então, por (72), que

$$\det(B) = \sum_{j=1}^n (-1)^{j-1} b_{1,j} \det(\hat{B}_j) = \sum_{j=1}^n (-1)^{j-1} b_{1,j} \cdot (-1) \det(\hat{A}_j)$$

que é igual a $-\det(A)$, provando, assim, o primeiro caso.

Para o segundo caso, suporemos que as linhas a serem trocadas são a primeira e a segunda. A demonstração deste caso é mais complicada, porque exige que façamos a expansão em cofatores dos determinantes de \hat{A}_j para cada $1 \leq j \leq n$. Denotando por $\hat{A}_{j,k}$ o k -ésimo cofator de \hat{A}_j , temos de (72) que

$$(74) \quad \det(\hat{A}_j) = \sum_{k=1}^{j-1} (-1)^{k-1} a_{2,k} \det(\hat{A}_{j,k}) + \sum_{k=j+1}^n (-1)^k a_{2,k} \det(\hat{A}_{j,k}).$$

Note que escrevemos $\det(\hat{A}_j)$ como a soma de dois somatórios, porque o cofator \hat{A}_j não contém nenhum elemento da j -ésima coluna de A . Pela mesma razão tivemos que reajustar o sinal das colunas no segundo somatório. Substituindo (74) na expressão de $\det(A)$, obtemos

$$\det(A) = \sum_{j=1}^n (-1)^{j-1} a_{1,j} \left(\sum_{k=1}^{j-1} (-1)^{k-1} a_{2,k} \det(\hat{A}_{j,k}) + \sum_{k=j+1}^n (-1)^k a_{2,k} \det(\hat{A}_{j,k}) \right).$$

Mas, pela distributividade da multiplicação de matrizes

$$(75) \quad \det(A) = \sum_{j=1}^n \left(\sum_{k=1}^{j-1} (-1)^{k+j-2} a_{1,j} a_{2,k} \det(\hat{A}_{j,k}) + \sum_{k=j+1}^n (-1)^{k+j-1} a_{1,j} a_{2,k} \det(\hat{A}_{j,k}) \right),$$

Vejamos o que aconteceria se aplicássemos a mesma fórmula à matriz B , obtida trocando-se a primeira e a segunda linhas de A . Denotando por $b_{i,j}$ a entrada na linha i e coluna j de B , temos que

$$b_{1,j} = a_{2,j} \quad \text{e} \quad b_{2,j} = a_{1,j}$$

para todo $1 \leq j \leq n$. Além disso,

$$\hat{B}_{j,k} = \hat{A}_{j,k},$$

quaisquer que sejam $1 \leq j, k \leq n$, porque nestes cofatores não aparecem entradas nas duas primeiras linhas de A , nem de B , e a troca de linhas não afeta as posições das colunas. Portanto, a fórmula para $\det(B)$ equivalente a (75) é exatamente a mesma para $\det(A)$, provando, assim, que $\det(A) = \det(B)$.

Resta-nos considerar o terceiro caso, que é aquele em que a primeira linha é trocada pela k -ésima linha de A , quando $k > 2$. Mas este caso segue-se, facilmente, dos outros dois. Para ver porque isto é verdade, começamos trocando a k -ésima linha de A com sua

segunda linha. Denotando por A' a matriz resultante desta troca temos, pelo primeiro caso, que

$$\det(A') = -\det(A).$$

Em seguida trocamos a primeira e a segunda linhas de A' , obtendo a matriz A'' que, pelo segundo caso, satisfaz

$$\det(A'') = -\det(A') = \det(A).$$

Finalmente, trocamos a segunda linha de A'' com a k -ésima linha de A'' . Denotando por A''' esta última matriz, uma nova aplicação do primeiro caso mostra que

$$-\det(A''') = -\det(A).$$

Isto prova o terceiro caso porque A''' é igual à matriz obtida trocando-se a primeira com a k -ésima linhas de A .

Há várias propriedades do determinante que são consequências diretas da troca de linhas; a primeira delas é a propriedade enunciada, em um caso especial, no lema 1.1.

PROPRIEDADE 1.3. *Se uma linha de uma matriz é multiplicada por uma constante, então seu determinante é multiplicado pela mesma constante.*

Se a linha em questão for a primeira, este resultado coincide com o que já foi provado no artigo 1.1. Caso contrário, trocamos a linha que foi multiplicada pela constante com primeira linha, aplicamos o lema 1.1 a esta nova matriz e, finalmente, destrocamos as linhas voltando à matriz inicial. Como estamos fazendo exatamente duas trocas de linha o sinal do determinante não é alterado, completando assim a demonstração desta propriedade.

PROPRIEDADE 1.4. *Se uma matriz quadrada tem duas linhas iguais, seu determinante é igual a zero.*

Se a j -ésima e k -ésima linhas de A coincidem, esta matriz não muda quando estas linhas forem trocadas. Logo, pela propriedade 1.2, $\det(A) = -\det(A)$, o que só pode ocorrer quando $\det(A) = 0$. Para provar a próxima propriedade basta mover a linha nula para o lugar da primeira e aplicar a expansão de cofatores.

PROPRIEDADE 1.5. *Se uma matriz quadrada tem uma linha nula, seu determinante é igual a zero.*

A importância da próxima propriedade está em que nos permite calcular facilmente o determinante de uma matriz triangular superior. Como, no próximo artigo, mostramos que o determinante de uma matriz coincide com o de sua forma escada; este resultado acaba sendo a chave para um método eficaz de calcular determinantes de matrizes cujas entradas são números. Antes, porém, precisamos de um resultado auxiliar.

LEMA 1.6. *Se a primeira coluna de uma matriz quadrada é nula, seu determinante é igual a zero.*

Vamos argumentar por indução no tamanho n da matriz. Se $n = 1$, a matriz é nula e o resultado é óbvio. Suponhamos que o resultado vale para toda matriz $(n - 1) \times (n - 1)$ e seja

$$A = \begin{bmatrix} 0 & a_{1,2} & \cdots & a_{1,n} \\ 0 & a_{2,2} & \cdots & a_{2,n} \\ 0 & a_{3,2} & \cdots & a_{3,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n,2} & \cdots & a_{n,n} \end{bmatrix}.$$

Aplicando a expansão por cofatores a esta matriz, obtemos

$$(76) \quad \det(A) = \sum_{j=2}^n (-1)^{j-1} a_{1,j} \det(\widehat{A}_j).$$

Note que a soma começa em $j = 2$, porque a entrada 1, 1 de A é igual a zero. Por outro lado, como as posições na primeira coluna de A sobrevivem em

$$\widehat{A}_2, \dots, \widehat{A}_n,$$

estes cofatores são matrizes $(n - 1) \times (n - 1)$ cuja primeira coluna é nula. Logo, pela hipótese de indução,

$$\det(\widehat{A}_2) = \cdots = \det(\widehat{A}_n) = 0.$$

Mas, substituindo isto em (76), obtemos $\det(A) = 0$, como queríamos mostrar.

PROPRIEDADE 1.7. *O determinante de uma matriz triangular superior é igual ao produto das entradas de sua diagonal.*

Também esta propriedade será provada por indução no tamanho n da matriz. Quando $n = 1$ a matriz é igual à única posição de sua diagonal, de modo que o resultado claramente vale neste caso. Suponhamos que $n \geq 2$ e que a propriedade é válida para toda matriz triangular superior de tamanho $(n - 1) \times (n - 1)$, e seja

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n-1} & a_{1,n} \\ 0 & a_{2,2} & \cdots & a_{2,n-1} & a_{2,n} \\ 0 & 0 & \cdots & a_{3,n-1} & a_{3,n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & a_{n,n} \end{bmatrix}.$$

Aplicando a A a expansão em (72), obtemos

$$(77) \quad \det(A) = a_{1,1} \det \det(\widehat{A}_1) + \sum_{j=2}^n (-1)^{j-1} a_{1,j} \det(\widehat{A}_j).$$

Como a primeira coluna dos cofatores

$$\widehat{A}_2, \dots, \widehat{A}_n$$

são iguais à primeira coluna de A com a primeira posição removida, segue-se do lema 1.6 que todos estes cofatores têm determinante igual a zero. Logo, por (77),

$$\det(A) = a_{1,1} \det \det(\widehat{A}_1)$$

Mas,

$$\widehat{A}_1 = \begin{bmatrix} a_{2,2} & \cdots & a_{2,n-1} & a_{2,n} \\ 0 & \cdots & a_{3,n-1} & a_{3,n} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & a_{n,n} \end{bmatrix}$$

é uma matriz triangular superior de tamanho $(n-1) \times (n-1)$, de modo que, pela hipótese de indução,

$$\det(\widehat{A}_1) = a_{2,2} \cdots a_{n,n}.$$

Portanto,

$$\det(A) = a_{1,1} \det(\widehat{A}_1) = a_{1,1} a_{2,2} \cdots a_{n,n},$$

como precisávamos mostrar.

1.3. Operações elementares por linha. A próxima propriedade dos determinantes diz respeito ao que ocorre quando aplicamos uma operação elementar por linha a uma matriz.

PROPRIEDADE 1.8. *O determinante de uma matriz não muda quando uma de suas linhas é substituída por ela mesma mais um múltiplo de outra linha.*

Começamos com o caso especial em que a operação é aplicada às duas primeiras linhas da matriz. Seja A uma matriz $n \times n$ como em (73) e seja B a matriz obtida substituindo-se a primeira linha de A por ela mesma mais α vezes a segunda linha, em que α é um número real qualquer. Então,

$$B = \begin{bmatrix} a_{1,1} + \alpha a_{2,1} & a_{1,2} + \alpha a_{2,2} & \cdots & a_{1,n} + \alpha a_{2,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ a_{3,1} & a_{3,2} & \cdots & a_{3,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{bmatrix}.$$

Expandindo o determinante de B em cofatores, obtemos por (72),

$$\det(B) = \sum_{j=1}^n (-1)^{j-1} (a_{1,j} + \alpha a_{2,j}) \det(\widehat{A}_j),$$

porque os cofatores de A e B são os mesmos. Pela distributividade da multiplicação,

$$\det(B) = \sum_{j=1}^n (-1)^{j-1} a_{1,j} \det(\hat{A}_j) + \alpha \sum_{j=1}^n (-1)^{j-1} a_{2,j} \det(\hat{A}_j).$$

Mas o primeiro somatório é igual a $\det(A)$ e o segundo ao determinante da matriz obtida substituindo-se a primeira linha de A pela segunda. Como este último determinante é igual a zero pela propriedade 1.4, temos que $\det(B) = \det(A)$. Finalmente precisamos mostrar que o caso geral pode ser reduzido ao caso especial que acabamos de provar. Digamos que a operação por linha consiste em substituir a j -ésima linha de A por ela mesma mais α vezes a i -ésima linha. Usando a notação do capítulo 4, podemos fazer esta operação executando a seguinte sequência de operações elementares

$$A \xrightarrow{T_{1,j}} A_1 \xrightarrow{T_{2,i}} A_2 \xrightarrow{L_{2,1}(\alpha)} A_3 \xrightarrow{T_{1,j}} A_4 \xrightarrow{T_{2,i}} A_5.$$

As duas primeiras trocas de linhas substituem a j -ésima linha de A pela primeira e sua i -ésima linha pela segunda. Portanto, pelo caso especial,

$$\det(A_3) = \det(A_2).$$

Como o número de trocas de linhas necessários para transformar A em A_2 é o mesmo usado para transformar A_3 em A_5 , a propriedade 1.2 nos permite concluir que

$$\det(A) = \det(A_5),$$

Como queríamos mostrar. Esta propriedade nos dá uma maneira simples de calcular determinantes usando eliminação.

PROPRIEDADE 1.9. *O determinante de uma matriz quadrada é igual ao produto das entradas na diagonal de sua forma escada.*

Pela propriedade 1.8 o determinante de uma matriz quadrada coincide com o de sua forma escada. Mas, pela propriedade 1.7, o determinante da forma escada é o produto das suas entradas diagonais.

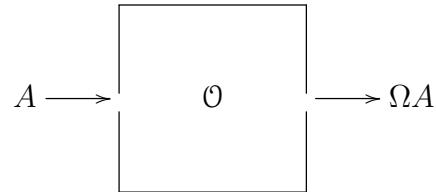
2. Decomposição de matrizes

Talvez devêssemos classificar a versão do método de eliminação gaussiana estudado na seção anterior como clássica, em oposição à versão moderna, que estudaremos nesta seção. Ambas são versões matriciais do método de adição, generalizado para sistemas com mais de duas incógnitas. Entretanto, na versão tradicional, as matrizes são apenas uma maneira conveniente de carregar o mínimo de informação possível, mantendo os coeficientes na posição correta, mas eliminando as variáveis e os símbolos para as operações de soma e multiplicação. A versão moderna, ao contrário, é intrinsecamente matricial e não faz sentido se não adotarmos esta linguagem, porque seu resultado é uma decomposição da matriz dada como um produto de duas outras matrizes, uma delas triangular superior, a

outra triangular inferior. Para chegar a isto começamos investigando como as operações elementares por linha podem ser descritas em termos de produtos de matrizes.

2.1. Matrizes elementares. Nossa ponto de partida é exatamente o pressuposto de que operações elementares por linha podem ser descritas em termos de produtos de matrizes. Supondo que isto seja verdade, tentaremos descobrir que matriz deveria efetuar esta operação. Uma vez identificada esta matriz, uma mera multiplicação de matrizes basta para provar que o pressuposto está correto.

Considerando um contexto um pouco mais geral, digamos que \mathcal{O} é um *oráculo*; isto é, um algoritmo que pode ser aplicado a qualquer matriz $n \times n$, mas cujo funcionamento interno desconhecemos. Imagine, agora, que descobrimos, ou fomos informados, de que \mathcal{O} apenas multiplica a matriz de entrada por uma matriz desconhecida Ω .



Neste caso, decifrar o oráculo resume-se a identificar Ω . Mas isto é muito fácil de fazer: basta dar como entrada a \mathcal{O} a matriz identidade $n \times n$. Se \mathcal{O} atua multiplicando por Ω a matriz que lhe foi dada como entrada, a saída neste caso será a própria Ω .

Passando ao caso que nos interessa, sejam $1 \leq i < j \leq n$ inteiros e suponhamos que \mathcal{O} é o oráculo que, ao receber uma matriz A , retorna a matriz obtida somando à linha j de A o produto de sua linha i por um escalar r . Levando em conta que a única posição não nula da linha i da matriz identidade I ocorre na diagonal e vale 1, podemos concluir que ao receber I este oráculo retornará a matriz

$$\begin{matrix}
 & & & i \\
 & & & \downarrow \\
 & & 1 & \\
 & & \ddots & \\
 & & 1 & r \\
 & & & \ddots \\
 & & & 1 \\
 j & & & & & & 1
 \end{matrix}$$

em que o r está localizado na posição ji e os espaços vazios estão todos preenchidos com zeros. No artigo 1.2 do capítulo 3 denotamos esta matriz por $L_{ji}(r)$. Portanto, podemos afirmar que:

se \mathcal{O} é implementado como a multiplicação da matriz dada na entrada por alguma matriz Ω , então $\Omega = L_{ji}(r)$.

Resta-nos verificar que $L_{ji}(r)A$ de fato é igual à matriz obtida a partir de A substituindo-se sua j -ésima linha por ela própria somada a r vezes sua i -ésima linha. Para isso, convém reescrever $L_{ji}(r)$ na forma

$$L_{ji}(r) = I + rE_{ji},$$

em que E_{ji} é a matriz $n \times n$ que tem todas as posições nulas, exceto a posição ji , que vale 1. Temos, então, pela distributividade da multiplicação de matrizes, que

$$L_{ji}(r)A = A + rE_{ji}A.$$

Como a única posição não nula de E_{ji} está na coluna i ,

$$E_{j,i}A = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ a_{j,1} & \cdots & a_{j,n} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \leftarrow \text{linha } i$$

Logo, a i -ésima linha de $E_{j,i}A$ é igual a j -ésima linha de A , ao passo que suas outras linhas são todas nulas. Portanto,

$$L_{ji}(r)A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{i,1} & \cdots & a_{i,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{bmatrix} + \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ a_{j,1} & \cdots & a_{j,n} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \leftarrow \text{linha } i$$

onde

$$L_{ji}(r)A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{i,1} + ra_{j,1} & \cdots & a_{i,n} + ra_{j,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{bmatrix} \leftarrow \text{linha } i$$

Isto justifica que chamemos de *elementares* às matrizes $L_{ij}(r)$, já que implementam as operações elementares por linha.

Por exemplo, o sistema (69) da página 97 tem

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 3 & 1 & 5 & 5 \\ 1 & 7 & -1 & 2 & 3 \\ 5 & -1 & -3 & 1 & 7 \end{bmatrix}$$

por matriz aumentada. Para pôr este sistema em forma triangular superior, utilizamos as operações elementares por linha listadas na tabela 1. Para cada uma destas operações escrevemos a matriz elementar correspondente, segundo a receita que acabamos de descobrir.

Operação por linha	matriz elementar
segunda linha menos o dobro da primeira	$L_{21}(-2)$
terceira linha menos a primeira	$L_{31}(-1)$
quarta linha menos o quíntuplo da primeira	$L_{41}(-5)$
terceira linha menos o sétuplo segunda	$L_{32}(-6)$
quarta linha mais o sétuplo segunda	$L_{42}(6)$
quarta linha mais sete meios da terceira	$L_{43}(7/2)$

TABELA 1. Operações por linha

Portanto,

$$(78) \quad L_{43}(7/2) \cdot L_{42}(6) \cdot L_{32}(-6) \cdot L_{41}(-5) \cdot L_{31}(-1) \cdot L_{21}(-2) \cdot A$$

deve ser igual à matriz triangular superior

$$U = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & 4 & -17 \\ 0 & 0 & 0 & -91/2 \end{bmatrix},$$

como é fácil de verificar efetuando as contas. Fácil, mas muito trabalhoso, a não ser que você peça ajuda a um computador.

As matrizes elementares têm três propriedades importantes que precisamos mencionar porque serão utilizadas mais adiante. A primeira, e mais óbvia, é que qualquer matriz elementar é triangular superior ou triangular inferior. Afinal, uma matriz elementar tem apenas uma posição fora da diagonal. Quando esta posição está acima da diagonal, temos uma matriz triangular superior; quando abaixo, temos uma matriz triangular inferior. Mais precisamente, se $1 \leq i, j \leq n$ são inteiros e $a \in \mathbb{R}$, então a matriz elementar $L_{ij}(a)$ de tamanho $n \times n$ é *triangular inferior* quando $i > j$ e *triangular superior* quando $i < j$. A segunda propriedade, muito mais interessante, afirma que $L_{ij}(a)$ é uma matriz inversível

quaisquer que sejam $i \neq j$ e $a \in \mathbb{R}$. Para provar isto, basta exibir um inverso para $L_{ij}(a)$, o que é muito fácil, porque o inverso desta matriz é $L_{ij}(-a)$. Para provar isto, lembre-se que, por definição

$$L_{ij}(a) = I + aE_{ij}, \text{ ao passo que } L_{ij}(-a) = I - aE_{ij}.$$

Portanto, pelas propriedades da multiplicação de matrizes

$$L_{ij}(a) \cdot L_{ij}(-a) = (I + aE_{ij})(I - aE_{ij}) = I + aE_{ij} - aE_{ij} - a^2E_{ij}^2 = I - a^2E_{ij}^2.$$

Como E_{ij} tem apenas uma entrada não nula (na posição i, j), seu quadrado só pode ser diferente de zero se $i = j$, que nunca é o caso quando tratamos de matrizes elementares. Logo, $E_{ij}^2 = 0$; donde

$$L_{ij}(a) \cdot L_{ij}(-a) = I,$$

como devíamos provar. A terceira propriedade de $L_{ij}(a)$ diz respeito ao seu determinante. Como $L_{ij}(a)$ pode ser obtida aplicando a I a operação elementar que substitui sua j -ésima linha por a vezes sua i -ésima linha, temos pela propriedade 1.8 do determinante que

$$\det(L_{ij}(a)) = \det(I) = 1.$$

O lema abaixo resume estes três fatos, para referência futura.

LEMA 2.1. *Sejam i e j inteiros positivos distintos menores ou iguais a n e a um número real. A matriz elementar $L_{ij}(a) = I + aE_{ij}$ tem determinante igual a 1 e sua inversa é igual a $L_{ij}(-a)$.*

No início desta seção mencionamos que as matrizes elementares nos permitem interpretar o resultado da eliminação gaussiana como uma decomposição de matrizes. Vamos ilustrar como isto pode ser feito usando o exemplo acima. Começamos multiplicando os dois lados da equação 78, à esquerda, por $L_{43}(-7/2)$, o que nos dá

$$(79) \quad L_{42}(6) \cdot L_{32}(-6) \cdot L_{41}(-5) \cdot L_{31}(-1) \cdot L_{21}(-2) \cdot A = L_{43}(7/2) \cdot U,$$

pois pelo lema 2.1, $L_{43}(-7/2)$ é a inversa de $L_{43}(7/2)$. Repetindo este procedimento para cada uma das matrizes que aparecem do lado esquerdo de (79), obtemos

$$A = L_{21}(-2) \cdot L_{31}(1) \cdot L_{41}(5) \cdot L_{32}(6) \cdot L_{42}(-6) \cdot L_{43}(7/2) \cdot U.$$

Seja

$$L = L_{21}(-2) \cdot L_{31}(1) \cdot L_{41}(5) \cdot L_{32}(6) \cdot L_{42}(-6) \cdot L_{43}(7/2).$$

Como o produto de matrizes triangulares inferiores é uma matriz triangular inferior, mostramos, que $A = LU$ é uma decomposição de A como o produto da matriz triangular inferior L , pela matriz triangular superior U . Esta maneira de interpretar a eliminação gaussiana, conhecida como *decomposição LU*, foi originalmente proposta por Tadeusz Banachiewicz em 1938 e redescoberta, independentemente, por P. D. Crout em 1941 e por Alan Turing em 1948.

2.2. Inversão de matrizes. Seja A a matriz quadrada $n \times n$ que desejamos inverter. Aplicando operações elementares por linha a A , obtemos sua forma escada U . Porém, como vimos no artigo 2.1, estas operações podem ser realizadas multiplicando A por uma sucessivas matrizes elementares, digamos L_1, \dots, L_r , de modo que

$$L_r \cdots L_1 A = U.$$

Mas, pela propriedade 1.9 do determinante, sabemos que

$$\det(A) = \det(U) = u_{1,1} \cdots u_{n,n},$$

em que $u_{1,1}, \dots, u_{n,n}$ são as entradas diagonais de U . Portanto, quando $\det(A) \neq 0$, a matriz U não tem nenhuma entrada nula em sua diagonal. Isto nos permite usar essas entradas para eliminar as posições de U acima da diagonal, começando da última coluna e recuando até chegar à primeira. O processo é análogo ao que fizemos para ir de A a U , só que agora está caminhando do fim da matriz para seu início e eliminando de baixo para cima. Por isso as matrizes elementares que usaremos terão a forma

$$L_{ji}(r) = I + rE_{ji},$$

só que, desta vez, com $j < i$. Por exemplo, para anular a entrada $u_{n-1,n}$ de U usamos a matriz

$$L_{n-1,n} \left(-\frac{u_{n-1,n}}{u_{n,n}} \right) = I - \left(\frac{u_{n-1,n}}{u_{n,n}} \right) E_{n-1,n},$$

Procedendo desta maneira encontramos matrizes elementares L_{r+1}, \dots, L_s , tais que

$$L_s \cdots L_{r+1} \cdot L_r \cdots L_1 A = \begin{bmatrix} u_{1,1} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & u_{n,n} \end{bmatrix}.$$

Denotando esta última matriz por D , temos que

$$D^{-1} = \begin{bmatrix} u_{1,1}^{-1} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & u_{n,n}^{-1} \end{bmatrix}$$

pois, como já vimos, as entradas diagonais de U não podem se anular. Logo,

$$(80) \quad D^{-1} \cdot L_s \cdots L_{r+1} \cdot L_r \cdots L_1 A = I;$$

em que I é a matriz identidade de mesma dimensão que A ; donde

$$(81) \quad A^{-1} = D^{-1} \cdot L_s \cdots L_{r+1} \cdot L_r \cdots L_1.$$

Para calcular efetivamente a inversa de A pela fórmula acima, precisamos da matriz diagonal D e das matrizes elementares usadas para reduzir A a D . Na prática, contudo, há uma maneira de sistematizar estes cálculos, que não requer o uso explícito das matrizes

elementares. Se M_1 e M_2 são matrizes $n \times n$, denotamos por $[M_1 | M_2]$ a matriz $n \times 2n$ obtida colando M_2 à esquerda de M_1 . Por exemplo, se

$$M_1 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \text{e} \quad M_2 = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

então

$$[M_1 | M_2] = \begin{bmatrix} a_{11} & a_{12} & b_{11} & b_{12} \\ a_{21} & a_{22} & b_{21} & b_{22} \end{bmatrix}.$$

A chave para a sistematização que desejamos é a observação de que se M_3 é uma terceira matriz $n \times n$, então

$$M_3 \cdot [M_1 | M_2] = [M_3 \cdot M_1 | M_3 \cdot M_2].$$

Em particular, temos de (80) e (81), que

$$D^{-1} \cdot L_s \cdots L_{r+1} \cdot L_r \cdots L_1 [A | I] = A^{-1}[A | I] = [I | A^{-1}].$$

Em outras palavras, aplicando a $[A | I]$ as mesmas operações por linha que reduzem A à matriz diagonal D e multiplicando a matriz resultante por D^{-1} , obtemos A^{-1} como o segundo bloco $n \times n$ de uma matriz cujo primeiro bloco é a identidade. Tomando mais uma vez

$$A = \begin{bmatrix} 1 & 4 & 6 & 1 \\ 2 & 8 & 5 & -1 \\ 3 & 25 & 2 & -5 \\ 4 & 2 & -1 & -3 \end{bmatrix},$$

como exemplo, construímos

$$[A | I] = \left[\begin{array}{cccc|cccc} 1 & 4 & 6 & 1 & 1 & 0 & 0 & 0 \\ 2 & 8 & 5 & -1 & 0 & 1 & 0 & 0 \\ 3 & 5 & 2 & -5 & 0 & 0 & 1 & 0 \\ 4 & 2 & -1 & -3 & 0 & 0 & 0 & 1 \end{array} \right]$$

cuja forma escada é

$$[U | M] = \left[\begin{array}{cccc|cccc} 1 & 4 & 6 & 1 & 1 & 0 & 0 & 0 \\ 0 & -7 & -16 & -8 & -3 & 0 & 1 & 0 \\ 0 & 0 & -7 & -3 & -2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 6 & 0 & 1 & -2 & 1 \end{array} \right].$$

Aplicando a esta matriz a operação elementar em que a segunda linha é substituída pelo produto de $-3/2$ pela *terceira* linha, resulta

$$\left[\begin{array}{cccc|cccc} 1 & 4 & 6 & 1 & 1 & 0 & 0 & 0 \\ 0 & -7 & -16 & -8 & -3 & 0 & 1 & 0 \\ 0 & 0 & -7 & 0 & -2 & \frac{3}{2} & -1 & \frac{1}{2} \\ 0 & 0 & 0 & 6 & 0 & 1 & -2 & 1 \end{array} \right].$$

Mais duas operações elementares elementares com base na quarta linha e chegamos à matriz

$$\left[\begin{array}{cccc|cccc} 1 & 4 & 6 & 0 & 1 & -\frac{1}{6} & \frac{1}{3} & -\frac{1}{6} \\ 0 & -7 & -16 & 0 & -3 & \frac{4}{3} & -\frac{5}{3} & \frac{4}{3} \\ 0 & 0 & -7 & 0 & -2 & \frac{3}{2} & -1 & \frac{1}{2} \\ 0 & 0 & 0 & 6 & 0 & 1 & -2 & 1 \end{array} \right].$$

Tomando, agora, como pivô a entrada -7 na posição 3, 3, eliminamos as posições acima dela na mesma coluna, do que resta a matriz

$$\left[\begin{array}{cccc|cccc} 1 & 4 & 0 & 0 & -\frac{5}{7} & \frac{47}{42} & -\frac{11}{21} & \frac{11}{42} \\ 0 & -7 & 0 & 0 & \frac{11}{7} & -\frac{44}{21} & \frac{13}{21} & \frac{4}{21} \\ 0 & 0 & -7 & 0 & -2 & \frac{3}{2} & -1 & \frac{1}{2} \\ 0 & 0 & 0 & 6 & 0 & 1 & -2 & 1 \end{array} \right].$$

Finalmente, usando a posição em 2, 2 como pivô eliminamos a entrada imediatamente acima dela, obtendo a matriz

$$\left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & \frac{9}{49} & -\frac{23}{294} & -\frac{25}{147} & \frac{109}{294} \\ 0 & -7 & 0 & 0 & \frac{11}{7} & -\frac{44}{21} & \frac{13}{21} & \frac{4}{21} \\ 0 & 0 & -7 & 0 & -2 & \frac{3}{2} & -1 & \frac{1}{2} \\ 0 & 0 & 0 & 6 & 0 & 1 & -2 & 1 \end{array} \right].$$

Portanto, pelo que vimos no início deste artigo,

$$A^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -7 & 0 & 0 \\ 0 & 0 & -7 & 0 \\ 0 & 0 & 0 & 6 \end{bmatrix} \cdot \begin{bmatrix} \frac{9}{49} & -\frac{23}{294} & -\frac{25}{147} & \frac{109}{294} \\ \frac{11}{7} & -\frac{44}{21} & \frac{13}{21} & \frac{4}{21} \\ -2 & \frac{3}{2} & -1 & \frac{1}{2} \\ 0 & 1 & -2 & 1 \end{bmatrix} = \begin{bmatrix} \frac{9}{49} & -\frac{23}{294} & -\frac{25}{147} & \frac{109}{294} \\ -\frac{11}{7 \cdot 7} & \frac{44}{7 \cdot 21} & -\frac{13}{7 \cdot 21} & -\frac{4}{7 \cdot 21} \\ \frac{2}{7} & -\frac{3}{7 \cdot 2} & \frac{1}{7} & -\frac{1}{7 \cdot 2} \\ 0 & \frac{1}{6} & -\frac{2}{6} & \frac{1}{6} \end{bmatrix}$$

é a inversa da matriz A .

2.3. Determinante do produto. Sejam A uma matriz quadrada de tamanho $n \times n$ e L_1, \dots, L_s as matrizes elementares que reduzem A à forma diagonal

$$D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix};$$

assim,

$$(82) \quad L_s \cdots L_1 \cdot A = D.$$

Se B for outra matriz, de mesmo tamanho que A , então

$$(83) \quad L_s \cdots L_1 \cdot AB = DB.$$

Logo, aplicando a propriedade 1.8 do determinante em (82) e (83), obtemos

$$\det(A) = \det(D) \quad \text{e} \quad \det(AB) = \det(DB).$$

Por outro lado, como, para cada $k = 1, \dots, n$, a k -ésima linha de DB é igual à k -ésima linha de B , multiplicada por λ_k , temos, pela propriedade 1.3, que

$$\det(A) = \lambda_1 \cdots \lambda_n \quad \text{e que} \quad \det(DB) = \lambda_1 \cdots \lambda_n \det(B).$$

Combinando as três últimas equações,

$$\det(AB) = \lambda_1 \cdots \lambda_n \det(B) = \det(A) \det(B).$$

Com isso mostramos a seguinte propriedade dos determinantes.

PROPRIEDADE 2.2. *Se A e B são matrizes quadradas de mesmo tamanho, então $\det(AB) = \det(A) \det(B)$.*

Frequentemente nos referiremos a esta propriedade, abreviadamente, pela expressão *o determinante de um produto é o produto dos determinantes*. Como consequência, temos as seguintes propriedades das matrizes inversíveis.

PROPOSIÇÃO 2.3. *Se A e B são matrizes quadradas de mesmo tamanho, então:*

- (a) *A é inversível se, e somente se, $\det(A) \neq 0$;*
- (b) *se A é inversível, então $\det(A^{-1}) = 1/\det(A)$;*
- (c) *AB é inversível se, e somente se, A e B são ambas inversíveis.*

DEMONSTRAÇÃO. A demonstração de que uma matriz de determinante não nulo tem inversa foi feita no artigo 2.2, quando mostramos como inverter uma matriz usando operações elementares por linha. Para provar a recíproca, suponhamos que A é inversível. Isto significa, por definição, que existe uma matriz A^{-1} tal que

$$A \cdot A^{-1} = I.$$

Logo, pela propriedade 2.2,

$$\det(A) \det(A^{-1}) = \det(I) = 1.$$

Portanto, se A é inversível, então, $\det(A) \neq 0$ e

$$\det(A^{-1}) = \frac{1}{\det(A)},$$

da qual decorrem, não somente a implicação que faltava em (a), como também a fórmula em (b). Finalmente, como

$$\det(AB) = \det(A) \det(B),$$

então $\det(AB) \neq 0$ se, e somente se, $\det(A) \neq 0$ e $\det(B) \neq 0$. O resultado em (c) segue, então, por (a). \square

Combinando a técnica de redução a uma matriz diagonal com a propriedade 2.2, podemos determinar, sem muito esforço, uma fórmula para o determinante da transposta de uma matriz. Seja A uma matriz quadrada. Transpondo as matrizes na igualdade (82), temos que

$$(84) \quad A^t \cdot L_1^t \cdots L_s^t = D^t = D,$$

pois D é uma matriz diagonal. Contudo, a transposta de uma matriz elementar, também é uma matriz elementar, pois

$$(I + aE_{i,j})^t = I^t + aE_{i,j}^t = I + aE_{j,i},$$

quaisquer que sejam $i \neq j$. Como as matrizes L_1, \dots, L_s são todas elementares, concluímos, pela a propriedade 2.2, que

$$\det(L_1^t \cdots L_s^t) = \det(L_1^t) \cdots \det(L_s^t) = \det(L_1) \cdots \det(L_s) = 1.$$

Levando isto em conta e aplicando, novamente, a propriedade 2.2, desta vez à igualdade em (84), obtemos

$$\det(A^t) = \det(A^t) \det(L_1^t \cdots L_s^t) = \det(D) = \det(A),$$

que é a igualdade que queríamos mostrar. Encerramos, enunciando este fato como uma propriedade do determinante, para referência futura.

PROPRIEDADE 2.4. *O determinante de uma matriz é igual ao de sua transposta.*

Exercícios

1. Use método de eliminação para calcular o determinante de cada uma das matrizes 3×3 abaixo.

$$(a) \begin{bmatrix} 2 & 5 & -1 \\ 4 & -1 & 2 \\ 6 & 4 & 0 \end{bmatrix}$$

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

$$(c) \begin{bmatrix} 4 & -1 & -3 \\ 3 & -2 & 5 \\ 2 & 3 & 4 \end{bmatrix}$$

$$(d) \begin{bmatrix} 3 & 1 & 2 \\ 2 & 0 & 1 \\ 4 & 3 & 4 \end{bmatrix}$$

$$(e) \begin{bmatrix} 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & -1 & 1 \end{bmatrix}$$

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

2. Use método de eliminação para calcular o determinante de cada uma das matrizes 4×4 abaixo.

$$(a) \begin{bmatrix} 1 & 3 & 0 & 1 \\ 1 & 2 & 1 & 1 \\ 2 & 4 & 2 & 1 \\ 1 & 0 & 2 & 0 \end{bmatrix}$$

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

$$(c) \begin{bmatrix} 1 & 2 & -3 & 5 \\ 0 & 1 & 4 & 3 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

3. Calcule a inversa da matriz resultante do seguinte produto de matrizes elementares:

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 6 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 9 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 8 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 5 & 1 \end{bmatrix}$$

4. Sejam v_1, \dots, v_n matrizes linha $1 \times n$ e $\delta(v_1, \dots, v_n)$ o determinante da matriz $n \times n$

$$\begin{bmatrix} - & v_1 & - \\ & \vdots & \\ - & v_n & - \end{bmatrix}$$

cujas linhas são iguais a v_1, \dots, v_n . Mostre que se v'_1 é uma matriz linha $1 \times n$ e $\lambda, \lambda' \in \mathbb{R}$, então

$$\delta(\lambda v_1 + \lambda' v'_1, \dots, v_n) = \lambda \delta(v_1, \dots, v_n) + \lambda' \delta(v'_1, \dots, v_n)$$

e que uma propriedade análoga vale para toda entrada $1 \leq k \leq n$ da função δ .

5. Prove que todas as propriedades do determinante nos artigos 1.2 e 1.3 continuam válidas quando trocamos a palavra *linha* pela palavra *coluna* em seus enunciados.

6. Em geral, um cofator de uma matriz A de tamanho $n \times n$ é uma matriz de tamanho $(n - 1) \times (n - 1)$, obtida removendo-se uma linha e uma coluna de A . Enuncie e prove uma fórmula análoga a (72) para a expansão de cofatores a partir da r -ésima linha da matriz.
Sugestão: use troca de linhas para reduzir o problema à expansão de cofatores pela primeira linha usada para definir determinante.
7. Enuncie e prove uma fórmula análoga à do exercício anterior para a expansão de cofatores a partir da r -ésima *coluna* da matriz.

Parte 3

Espaços vetoriais e transformações lineares

CAPÍTULO 6

Espaços multidimensionais

Neste capítulo introduzimos os espaços de dimensão n e suas operações, generalizando assim boa parte do que fizemos sobre o plano no capítulo 1. Além disso, investigamos o significado geométrico de seus subespaços.

1. Sistemas homogêneos

Se um sistema linear for impossível ou determinado, há pouco o que dizer sobre ele. No primeiro caso não há solução; no segundo caso, há uma única solução que pode ser encontrada combinando o método de eliminação com a resolução de um sistema triangular por substituição. O caso mais interessante ocorre quando o sistema é indeterminado, porque isto requer que sejamos capazes de expressar suas *infinitas* soluções de maneira que nos permita analisar seu comportamento. Este é o principal tema desta primeira seção.

1.1. Representação matricial de sistemas lineares. Começaremos investigando um exemplo. Considere o sistema linear

$$(85) \quad \begin{cases} x_1 - x_2 + x_3 - x_4 = 1 \\ x_2 - 3x_3 + 7x_4 = 2. \end{cases}$$

Como se trata de um sistema em forma escada, podemos resolvê-lo por substituição. Calculando o valor de x_2 a partir da segunda equação encontramos

$$x_2 = 2 + 3x_3 - 7x_4.$$

Substituindo na primeira equação, podemos escrever x_1 na forma

$$x_1 = 3 + 2x_3 - 6x_4.$$

Como x_3 e x_4 são parâmetros, obteremos uma solução do sistema se escolhermos para ambos o valor zero. Em outras palavras,

$$x_1 = 3, \quad x_2 = 2, \quad x_3 = 0 \quad \text{e} \quad x_4 = 0$$

é uma solução. Bem mais interessante é o que acontece quando substituímos

$$(86) \quad x_1 = 2x_3 - 6x_4 \quad \text{e} \quad x_2 = 3x_3 - 7x_4$$

no lado esquerdo das equações do sistema. Fazendo isso, na primeira equação, obtemos

$$(2x_3 - 6x_4) - (3x_3 - 7x_4) + x_3 - x_4 = 0$$

e na segunda,

$$(3x_3 - 7x_4) - 3x_3 + 7x_4 = 0.$$

Em outras palavras, as fórmulas em (86) determinam uma solução paramétrica do sistema

$$\begin{cases} x_1 - x_2 + x_3 - x_4 = 0 \\ x_2 - 3x_3 + 7x_4 = 0, \end{cases}$$

cujos termos constantes são nulos. Pensando sobre a maneira como as expressões para x_1 e x_2 foram obtidas, chegará rapidamente à conclusão de que seria de esperar que ambas as expressões acima viessem a se anular.

Para provar que o que verificamos no sistema (85) vale em geral, é conveniente introduzir a representação de um sistema linear como uma fórmula matricial. Para isso, seja

$$(87) \quad \begin{cases} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n = b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n = b_2 \\ \vdots \qquad \qquad \vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n = b_m \end{cases}$$

em que os coeficientes das variáveis x_1, \dots, x_n e os termos independentes são números reais. Se

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{bmatrix}, \quad X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \text{e} \quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

então o produto AX nos dá uma matriz coluna $m \times 1$ cujas linhas são as equações do sistema (87). Portanto, este sistema pode ser escrito, compactamente, na forma

$$(88) \quad AX = b.$$

Vamos sempre nos referir a A como *matriz do sistema*, a X como *matriz das variáveis* e a b como *matriz dos termos constantes*. Finalmente, a uma solução

$$x_1 = \alpha_1, x_2 = \alpha_2, \dots, x_n = \alpha_n,$$

do sistema (87) associamos a matriz coluna

$$v = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}.$$

Por exemplo, a matriz do sistema (85) é

$$\begin{bmatrix} 1 & -1 & 1 & -1 \\ 0 & 1 & -3 & 7 \end{bmatrix}$$

ao passo que suas matrizes de variáveis e de termos constantes são, respectivamente,

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \quad \text{e} \quad \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Já

$$\begin{bmatrix} 3 + 2x_3 - 6x_4 \\ 2 + 3x_3 - 7x_4 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 2 \\ 3 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -6 \\ -7 \\ 0 \\ 1 \end{bmatrix}$$

é a matriz correspondente à solução geral de (85). Nesta notação, o que descobrimos sobre as soluções deste sistema foi que

$$(89) \quad \begin{bmatrix} 2x_3 - 6x_4 \\ 3x_3 - 7x_4 \\ x_3 \\ x_4 \end{bmatrix}$$

é solução de

$$\begin{bmatrix} 1 & -1 & 1 & -1 \\ 0 & 1 & -3 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

A representação matricial de um sistema linear é especialmente adequada quando se trata de analisar suas soluções. Digamos que o sistema esteja escrito na forma (88) e que X_0 e X_1 sejam duas de suas soluções. Em outras palavras,

$$AX_0 = b \quad \text{e} \quad AX_1 = b.$$

Disto podemos concluir que

$$A(X_1 - X_0) = AX_1 - AX_0 = b - b = 0.$$

Logo, $X_1 - X_0$ é solução do sistema linear $AX = 0$, que chamaremos de *sistema homogêneo associado* a $AX = b$. Em geral, um sistema linear é *homogêneo* se os termos

constantes de suas equações são todos iguais a zero. Reciprocamente, se Y_0 é solução de $AX = 0$, então $X_0 + Y_0$ é solução de $AX = b$, pois

$$A(X_0 + Y_0) = AX_0 + AY_0 = b + 0 = b.$$

Portanto,

todas as soluções de um sistema linear podem ser obtidas somando-se a qualquer uma delas as soluções do sistema homogêneo associado.

Com isto, reproduzimos de maneira completamente geral o que havíamos visto no exemplo (85), porque (89) é, de fato, a diferença entre duas soluções daquele sistema.

Isto sugere que, a chave para entender como são construídas as soluções de um sistema indeterminado encontra-se no sistema homogêneo a ele associado.

1.2. Alguns exemplos de sistemas homogêneos.

As soluções do sistema

$$(90) \quad \begin{cases} x_1 - 2x_2 + 5x_3 = 0 \\ x_2 + 7x_3 = 0, \end{cases}$$

são dadas, em forma paramétrica, por

$$x_1 = -19x_3 \quad \text{e} \quad x_2 = -7x_3,$$

cuja representação matricial é

$$x_3 \begin{bmatrix} -19 \\ -7 \\ 1 \end{bmatrix}.$$

Logo, todas as soluções do sistema (90) são múltiplas de uma única solução.

Considere, agora, o sistema cuja única equação é

$$x_1 + x_2 + x_3 = 0.$$

Desta vez as soluções podem ser escritas em forma paramétrica como

$$x_1 = -x_2 - x_3,$$

de modo que sua representação matricial é

$$\begin{bmatrix} -x_2 - x_3 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

Copiando o que fizemos no capítulo 1, podemos dar uma interpretação geométrica a estas duas soluções. Para isso, definimos um vetor do espaço como um segmento orientado que tem uma de suas extremidades fixa da origem. Escolhemos, então, três vetores

unitários e dois-a-dois perpendiculares, que desempenharão o papel de base ortonormal do espaço. Isto nos permite associar a cada vetor v do espaço a tripla de números reais correspondente às projeções (com o devido sinal) de v sobre os vetores da base. Finalmente, como no caso do plano, identificamos cada tripla a uma matriz coluna 3×1 .

Com isso as soluções do sistema linear (90) consistem de todos os múltiplos do vetor $(-19, -7, 1)$. Como no caso do plano, isto representa uma reta que passa pela origem. Por outro lado, as soluções da equação $x_1 + x_2 + x_3 = 0$ correspondem a todos os vetores da forma

$$x_2(-1, 1, 0) + x_3(-1, 0, 1).$$

Como esses dois vetores não são múltiplos um do outro, o conjunto formado a partir deles não pode ser uma reta. Mas isto significa que há duas direções diferentes a partir das quais os vetores do conjunto solução de $x_1 + x_2 + x_3 = 0$ são construídos. Em outras palavras, esses vetores descrevem um plano dentro do espaço.

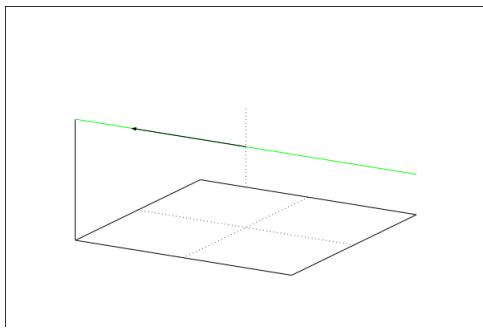


FIGURA 1. A reta de vetor diretor $(-19, -7, 1)$.

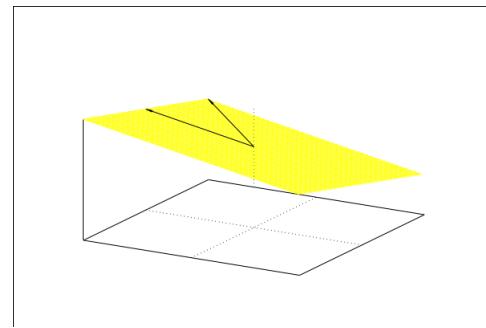


FIGURA 2. O plano $x_1 + x_2 + x_3 = 0$.

Para nosso terceiro exemplo, considere o sistema

$$\begin{cases} x + y + z + w &= 0 \\ x + 2y - 3z + w &= 0 \\ 5x + 7y - 3z + 5w &= 0, \end{cases}$$

cuja matriz é

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & -3 & 1 \\ 5 & 7 & -3 & 5 \end{bmatrix}$$

Aplicando eliminação gaussiana a A , obtemos a matriz

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

que corresponde ao sistema triangular

$$\begin{cases} x + y + z + w = 0, \\ y - 4z = 0. \end{cases}$$

Resolvendo este último sistema por substituição, podemos escrever suas soluções em forma paramétrica como

$$x = -5z - w \quad \text{e} \quad y = 4z,$$

em que z e w funcionam como parâmetros. Substituindo estes valores de x e y na matriz coluna formada pelas variáveis, verificamos que, quaisquer que sejam os valores escolhidos para z e w , a matriz

$$X_0 = \begin{bmatrix} -5z - w \\ 4z \\ z \\ w \end{bmatrix}$$

satisfaz $AX_0 = 0$, de modo que X_0 representa a solução do sistema quando este é expresso em forma matricial. Utilizando operações com matrizes, a mesma solução pode ser expressa na forma

$$X_0 = z \begin{bmatrix} -5 \\ 4 \\ 1 \\ 0 \end{bmatrix} + w \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Nossa interpretação geométrica para as soluções da equação $x_1 + x_2 + x_3 = 0$ sugere que, se pudéssemos considerar as matrizes coluna 4×1 como vetores em algum “espaço”, a solução deste sistema descreveria um plano. Foi exatamente este tipo de consideração que levou os matemáticos do século XIX a introduzir espaços de dimensão maior que três.

1.3. Um pouco de história. Em seu livro *Desenvolvimentos da Matemática no século XIX* [3, p. 156], Felix Klein diz que, tendo em vista as fórmulas para transformações no plano e no espaço

uma generalização parece quase óbvia: substituir as três variáveis x, y, z por qualquer número n de variáveis e, correspondentemente, estudar a geometria de um espaço de n dimensões. Esta ideia era tão natural que só podemos falar em um verdadeiro avanço tendo sido feito quando houve um interesse mais penetrante neste extenso domínio e a teoria necessária foi construída de maneira precisa.

Entre os primeiros matemáticos a estudarem de maneira sistemática a geometria em n -dimensões encontram-se George Green, Augustin-Louis Cauchy, Julius Plücker, Hermann Grassman e A. Cayley que, como já vimos, definiu as matrizes e suas operações básicas.

Como a citação de Klein sugere, a ideia de que há espaços de dimensão maior do que três vinha sendo considerada desde o século XVIII. No verbete *dimension* da *Encyclopédie* que publicou com Diderot a partir de 1751, D'Alembert escreve,

um homem astuto que conheço crê que podemos considerar a duração
como uma quarta dimensão.

Isto, claro, não passa de uma opinião. Lagrange, contudo, é muito mais explícito. Em sua *Teoria das funções analíticas*, publicada em 1797, ele escreve

Pode-se aplicar a teoria das funções na mecânica. Até agora as funções se referem basicamente ao tempo, que sempre designaremos por t ; e como a posição de um ponto no espaço depende de três coordenadas retangulares x , y e z , suporemos, nos problemas de mecânica, que estas coordenadas dependem do tempo. Assim, podemos considerar a mecânica como uma geometria em quatro dimensões e a análise mecânica como uma extensão da análise geométrica.

Na verdade, o uso generalizado de mais de quatro coordenadas teve seu prenúncio em outra obra de Lagrange, a *Mecânica Analítica*, que antecede a obra citada anteriormente em nove anos.

No século XX a ideia do tempo como quarta dimensão foi introduzido por H. Minkowski em 1907 como uma maneira de geometrizar a teoria da relatividade que Einstein havia publicado dois anos antes. Apesar de ter inicialmente reagido de maneira negativa à proposta de Minkowski, Einstein veio a entender a importância desta formulação geométrica e acabou por fazer dela a base sobre a qual construiu sua teoria da gravitação em 1915.

A discussão sobre o significado da quarta dimensão deu origem a inúmeros livros populares, o mais famoso dos quais é, provavelmente, *Flatland: A Romance of Many Dimensions*, publicado em 1884 por Edwin Abbott. Nele, figuras planas tentam entender o que significa a terceira dimensão a partir daquilo que podem ver, dada sua limitação de só enxergar o que está no plano que habitam. Da mesma forma nós, habitantes de um espaço tridimensional, podemos tentar imaginar como seria um objeto que só existe em um espaço de quatro dimensões a partir de suas projeções em três dimensões. Teremos oportunidade de fazer isto no artigo 1.3 do capítulo 7.

Hoje em dia toda esta conversa sobre quatro dimensões empalidece diante do que prometem os adeptos da teoria de cordas, cujo modelo do universo requer que habitemos um

espaço de 11 dimensões, 7 das quais são tão curvadas que não conseguimos identificá-las. Espaços de dimensões grandes também aparecem quando tentamos descrever a configuração dos planetas no sistema solar. Como três coordenadas são necessárias para determinar a posição de cada um dos oito planetas, a configuração de todo o sistema de planetas, sem contar asteróides e satélites, só pode ser feita em um espaço de $3 \times 8 = 24$ dimensões, em que cada grupo de 3 coordenadas representa a posição de um planeta.

2. O espaço \mathbb{R}^n

Seguindo o que aprendemos na seção anterior, introduziremos nesta seção os espaços de dimensão n como objetos geométricos e estudamos algumas de suas propriedades mais elementares.

2.1. O espaço \mathbb{R}^n . Consideraremos n -uplas de números reais como sendo vetores em um espaço de n dimensões (ou n -dimensional) que denotaremos por

$$\mathbb{R}^n = \underbrace{\mathbb{R} \times \cdots \times \mathbb{R}}_{n \text{ vezes}},$$

Como fizemos no caso do plano, estes vetores serão escritos, indiscriminadamente, como n -uplas ou matrizes coluna. Assim, conforme nossas necessidades, escreveremos

$$v = (a_1, \dots, a_n) \quad \text{ou} \quad v = \begin{bmatrix} a_1, \\ \vdots \\ a_n \end{bmatrix}$$

para representar um vetor do \mathbb{R}^n . Tais vetores podem ser somados e multiplicados por escalares, exatamente como já fizemos nos casos em que a dimensão era dois. Mais precisamente, se

$$(91) \quad v = (\alpha_1, \dots, \alpha_n) \quad \text{e} \quad u = (\beta_1, \dots, \beta_n)$$

então

$$v + u = (\alpha_1 + \beta_1, \dots, \alpha_n + \beta_n);$$

ao passo que

$$\lambda v = (\lambda\alpha_1, \dots, \lambda\alpha_n),$$

qualquer que seja o escalar $\lambda \in \mathbb{R}$. Naturalmente não há nenhuma novidade nisto, uma vez que estes vetores também podem ser considerados como matrizes. Pela mesma razão,

estas operações satisfazem as seguintes propriedades:

$$\begin{array}{ll} (u + v) + w = u + (v + w) & 0 \cdot u = 0 \text{ e } 1 \cdot u = u \\ u + v = v + u & \lambda(u + v) = \lambda u + \lambda v \\ u + 0 = u & (\lambda + \mu)u = \lambda u + \mu u \\ u + (-1)u = 0 & (\lambda\mu)u = \lambda(\mu u) \end{array}$$

em que $u, v, w \in \mathbb{R}^n$, $\lambda, \mu \in \mathbb{R}$ e 0 é o vetor $(0, \dots, 0)$. Não precisamos provar nenhuma destas propriedades aqui, uma vez que já as propriedades análogas para matrizes foram demonstradas no capítulo 3.

Quanto ao produto interno, como nada temos senão as coordenadas dos vetores, só nos resta defini-lo a partir destas coordenadas. Descrevendo as coordenadas de u e v como em (91), definimos

$$\langle u | v \rangle = \sum_{j=1}^n \alpha_i \beta_i,$$

que a notação matricial nos permite escrever de forma compacta como

$$(92) \quad \langle u | v \rangle = u^t v;$$

em que, como sempre, fazemos os vetores u e v corresponderem a matrizes coluna. Usando as propriedades das operações com matrizes, podemos facilmente provar que o produto interno assim definido satisfaz as mesmas propriedades que o produto interno usual do plano. Relembrando, o produto interno é

distributivo: $\langle u | v_1 + v_2 \rangle = \langle u | v_1 \rangle + \langle u | v_2 \rangle$;

homogêneo: $\langle v_1 | \lambda v_2 \rangle = \lambda \langle v_1 | v_2 \rangle$;

simétrico: $\langle v_1 | v_2 \rangle = \langle v_2 | v_1 \rangle$;

positivo definido: $\langle u | u \rangle \geq 0$ e $\langle u | u \rangle = 0$ se, e somente se, $u = 0$;

quaisquer que sejam os vetores u, v_1 e v_2 do \mathbb{R}^n e o escalar λ . Como nos casos do plano e do espaço, diremos que dois vetores são *ortogonais* se o produto interno deles é nulo. Também a *norma euclidiana*, ou *comprimento* de um vetor, pode ser definida a partir do produto interno por

$$\|u\| = \sqrt{\langle u | u \rangle};$$

de modo que

$$\|(\alpha_1, \dots, \alpha_n)\| = \sqrt{\alpha_1^2 + \dots + \alpha_n^2}.$$

Como no caso do plano, diremos que dois vetores são *ortogonais* se seu produto interno é igual a zero. No artigo 5.4 veremos como definir ângulos entre vetores do \mathbb{R}^n , o que nos permitirá justificar porque, mesmo no \mathbb{R}^n , dois vetores cujo produto interno é zero devem ser considerados como perpendiculares.

2.2. Geometria multidimensional. Como vimos no artigo 1.2, o conjunto solução da equação $x_1 + x_2 + x_3 = 0$ é um plano e pode ser descrito como o conjunto de todos os vetores da forma

$$a_1u_1 + a_2u_2 \quad \text{em que} \quad u_1 = (-1, 1, 0), \quad u_2 = (-1, 0, 1) \quad \text{e} \quad a_1, a_2 \in \mathbb{R}.$$

Em geral, se $v, w_1, \dots, w_m \in \mathbb{R}^n$, diremos que v é uma *combinação linear* de w_1, \dots, w_m se existem números reais a_1, \dots, a_m tais que

$$v = a_1w_1 + \dots + a_mw_m.$$

Denotaremos o conjunto dos vetores que são combinações lineares de w_1, \dots, w_m por

$$\langle w_1, \dots, w_m \rangle,$$

e diremos que este conjunto é *gerado* por w_1, \dots, w_m . Assim, o plano de equação $x_1 + x_2 + x_3 = 0$ corresponde a $\langle u_1, u_2 \rangle$. Com esta notação, podemos facilmente generalizar retas e planos para espaços multidimensionais. Se w_1 e w_2 são vetores não colineares do \mathbb{R}^n , diremos que $\langle w_1 \rangle$ é a *reta* cujo vetor diretor é w_1 e que $\langle w_1, w_2 \rangle$ é o *plano* gerado por w_1 e w_2 . Como seria de esperar, se

$$v = a_1w_1 + a_2w_2$$

é um vetor do plano $W = \langle w_1, w_2 \rangle$, então a reta $\langle v \rangle$ está contida no plano W , pois

$$bv = b(a_1w_1 + a_2w_2) = (ba_1)w_1 + (ba_2)w_2 \in W.$$

Quando o espaço ambiente é o \mathbb{R}^3 , o conjunto solução S de um sistema linear que tem uma única equação não nula

$$a_1x_1 + a_2x_2 + a_3x_3 = 0,$$

é sempre um plano. Para provar isto, suporemos que $a_1 \neq 0$, o que nos permite reescrever a equação anterior na forma

$$x_1 = -\frac{a_2}{a_1}x_2 - \frac{a_3}{a_1}x_3.$$

Assim, $(x_1, x_2, x_3) \in S$ se, e somente se,

$$(x_1, x_2, x_3) = \left(-\frac{a_2}{a_1}x_2 - \frac{a_3}{a_1}x_3, x_2, x_3 \right) = x_2 \frac{1}{a_1}(-a_2, a_1, 0) + x_3 \frac{1}{a_1}(-a_3, 0, a_1).$$

Logo,

$$S = \langle (-a_2, a_1, 0), (-a_3, 0, a_1) \rangle.$$

Este argumento pode ser facilmente adaptado ao caso em que $a_1 = 0$ e $a_2 \neq 0$ ou $a_3 \neq 0$. Reciprocamente, se W for plano gerado pelos vetores

$$w_1 = (\alpha_1, \alpha_2, \alpha_3) \quad \text{e} \quad w_2 = (\beta_1, \beta_2, \beta_3),$$

então

$$\mathbf{n} = (\alpha_2\beta_3 - \alpha_3\beta_2, -\alpha_1\beta_3 + \alpha_3\beta_1, \alpha_1\beta_2 - \alpha_2\beta_1)$$

é ortogonal a w_1 e w_2 , como você pode facilmente confirmar calculando os produtos internos entre \mathbf{n} e os vetores w_1 e w_2 . Logo, pelas propriedades do produto interno, qualquer vetor v que for combinação linear de w_1 e w_2 também será ortogonal a \mathbf{n} . Assim, se $v = (x_1, x_2, x_3)$, então

$$0 = \langle v | \mathbf{n} \rangle = (\alpha_2\beta_3 - \alpha_3\beta_2)x_1 + (-\alpha_1\beta_3 + \alpha_3\beta_1)x_2 + (\alpha_1\beta_2 - \alpha_2\beta_1)x_3$$

é a equação do plano W gerado por w_1 e w_2 . Por exemplo, o plano gerado por

$$w_1 = (1, 1, 1) \quad \text{e} \quad w_2 = (1, 0, -1)$$

é o conjunto solução da equação $-x_1 + 2x_2 - x_3 = 0$.

Uma consequência do que acabamos de ver é que dois planos diferentes do \mathbb{R}^3 sempre se intersectam em uma reta. Para entender porque isto é verdade, observe que se os planos são diferentes, suas equações não podem ser múltiplos constantes uma da outra. Mas isto significa que, depois da eliminação, teremos um sistema em forma escada com duas equações não nulas. Resolvendo o sistema, escreveremos duas das variáveis em função da restante. Logo, os vetores no conjunto solução deste sistema serão todos múltiplos de um único vetor, o que caracteriza uma reta.

No \mathbb{R}^4 , entretanto, as coisas são um pouco diferentes. A dimensão extra cria espaço suficiente para que a interseção de dois planos possa ser apenas um ponto. Por exemplo, as duas primeiras coordenadas de um vetor no plano $U_1 = \langle e_3, e_4 \rangle$ são nulas, ao passo que, no plano $U_2 = \langle e_1, e_2 \rangle$, as duas últimas coordenadas são nulas. Portanto, as quatro coordenadas de um vetor na interseção de U_1 e U_2 serão todas nulas, de modo que $U_1 \cap U_2 = \{(0, 0, 0, 0)\}$ contém um único ponto.

Voltando aos planos do \mathbb{R}^3 , vimos que podem ser expressos de duas maneiras diferentes: como o conjunto gerado por dois vetores que não são colineares, ou como o conjunto solução de um sistema linear homogêneo com uma única equação diferente de zero. É razoável perguntar se o mesmo vale no \mathbb{R}^4 . Dada a maneira como definimos planos do \mathbb{R}^n no início deste artigo, o que estamos perguntando é se o conjunto solução no \mathbb{R}^4 de um sistema linear homogêneo que tem apenas uma equação continua sendo um plano. Vejamos o que acontece quando H é o conjunto solução de

$$x_1 + x_2 + x_3 + x_4 = 0.$$

Procedendo como fizemos no caso do \mathbb{R}^3 , podemos afirmar que $(x_1, x_2, x_3, x_4) \in H$ se, e somente se,

$$x_1 = -x_2 - x_3 - x_4;$$

onde

$$H = \{(-x_2 - x_3 - x_4, x_2, x_3, x_4) \mid x_2, x_3, x_4 \in \mathbb{R}\}.$$

Contudo

$$(-x_2 - x_3 - x_4, x_2, x_3, x_4) = x_2(-1, 1, 0, 0) + x_3(-1, 0, 1, 0) + x_4(-1, 0, 0, 1);$$

de modo que, usando a notação introduzida acima, podemos escrever

$$H = \langle (-1, 1, 0, 0), (-1, 0, 1, 0), (-1, 0, 0, 1) \rangle.$$

Portanto, os vetores de H são combinações lineares de três vetores; e não de dois vetores, como esperaríamos se H fosse um plano. É claro que isto não exclui a possibilidade de haver dois vetores que gerem H . Mas isto só é possível se existirem dois vetores w_1 e w_2 tais que

$$(93) \quad (-1, 1, 0, 0) = a_1 w_1 + a_2 w_2$$

$$(94) \quad (-1, 0, 1, 0) = b_1 w_1 + b_2 w_2$$

$$(95) \quad (-1, 0, 0, 1) = c_1 w_1 + c_2 w_2,$$

para escolhas adequadas de números reais a_1, a_2, b_1, b_2, c_1 e c_2 . Contudo,

$$b_1(-1, 1, 0, 0) - a_1(-1, 0, 1, 0) = b_1(a_1 w_1 + a_2 w_2) - a_1(b_1 w_1 + b_2 w_2) = (b_1 a_2 - a_1 b_2) w_2.$$

Como

$$c_1(-1, 1, 0, 0) - a_1(-1, 0, 0, 1) = c_1(a_1 w_1 + a_2 w_2) - a_1(c_1 w_1 + c_2 w_2) = (c_1 a_2 - a_1 c_2) w_2;$$

concluímos que

$$b_1(-1, 1, 0, 0) - a_1(-1, 0, 1, 0) \quad \text{e} \quad c_1(-1, 1, 0, 0) - a_1(-1, 0, 0, 1)$$

são colineares; isto é, existe $r \in \mathbb{R}$ tal que

$$b_1(-1, 1, 0, 0) - a_1(-1, 0, 1, 0) = r(c_1(-1, 1, 0, 0) - a_1(-1, 0, 0, 1)).$$

Mas isto equivale a dizer que

$$0 = (b_1 - rc_1)(-1, 1, 0, 0) - a_1(-1, 0, 1, 0) + a_1(-1, 0, 0, 1) = (-b_1 + rc_1, b_1 - rc_1, a_1, a_1);$$

onde podemos concluir que

$$a_1 = b_1 - rc_1 = 0.$$

Substituindo $a_1 = 0$ em (93), vemos que

$$(96) \quad (-1, 1, 0, 0) = a_2 w_2.$$

Por outro lado, obtemos de (94) e (95), que

$$(-1, 0, 1, 0) - r(-1, 0, 0, 1) = (b_1 - rc_1)w_1 + (b_2 - rc_2)w_2 = (b_2 - rc_2)w_2,$$

pois já sabemos que $b_1 - rc_1 = 0$. Multiplicando esta última equação por a_2 , temos

$$a_2(-1, 0, 1, 0) - ra_2(-1, 0, 0, 1) = (b_2 - rc_2)a_2 w_2 = (b_2 - rc_2)(-1, 1, 0, 0),$$

em que a última igualdade segue de (96). Logo,

$$(-a_2 + ra_2, 0, a_2, -ra_2) = (b_2 - rc_2)(-1, 1, 0, 0),$$

Em particular, $a_2 = 0$, o que não é possível pois, como $a_1 = 0$, teríamos que

$$0 = a_1 w_1 + a_2 w_2 = (-1, 1, 0, 0)$$

que é uma contradição.

O argumento acima mostra que o conjunto H não é um plano, muito menos uma reta: para poder identificá-lo, introduziremos uma nova terminologia. Diremos que o conjunto solução de uma equação

$$(97) \quad a_1x_1 + \cdots + a_nx_n = 0$$

em n -variáveis, cujos coeficientes não são todos nulos, é um *hiperplano* do \mathbb{R}^n . Considerando $\mathbf{n} = (a_1, \dots, a_n) \neq 0$ como um vetor do \mathbb{R}^n , temos que

$$\langle v | \mathbf{n} \rangle = a_1x_1 + \cdots + a_nx_n,$$

qualquer que seja o vetor $v = (x_1, \dots, x_n)$ do \mathbb{R}^n . Portanto, um vetor v pertence ao hiperplano de equação (97) se, e somente se, $\langle v | \mathbf{n} \rangle = 0$. Em outras palavras, um hiperplano também pode ser definido como o conjunto dos vetores do \mathbb{R}^n que são ortogonais a um dado vetor não nulo.

3. Subespaços

Nesta seção estudaremos as propriedades do conjunto solução de um sistema homogêneo em toda a sua generalidade.

3.1. Propriedades dos sistemas homogêneos. Seja A uma matriz $m \times n$ e sejam v_1 e v_2 duas soluções do sistema homogêneo $AX = 0$. Em outras palavras, estamos supondo que

$$Av_1 = 0 \quad \text{e} \quad Av_2 = 0.$$

Mas disto podemos deduzir que

$$A(v_1 + v_2) = Av_1 + Av_2 = 0 + 0 = 0;$$

de modo que $v_1 + v_2$ também é solução do sistema $AX = 0$. Analogamente, se $Av = 0$ e λ é um número real, então

$$A(\lambda v) = \lambda Av = \lambda \cdot 0 = 0,$$

que nos permite concluir que λv também é solução do mesmo sistema. Resumindo, provamos que

1. a soma de duas soluções de um sistema homogêneo, e
2. o produto da solução de um sistema homogêneo por um escalar;

também são soluções do mesmo sistema. Nenhuma destas duas propriedades vale quando o sistema não é homogêneo. Contudo, estas propriedades são muito convenientes, porque permitem gerar novas soluções a partir de outras, já conhecidas. Esta é uma das razões pelas quais, em vez de analisarmos diretamente as soluções de um sistema indeterminado que não é homogêneo, preferimos passar ao sistema homogêneo a ele associado.

Seguindo uma tendência que vigora na matemática desde meados do século XX, interpretaremos essas propriedades das soluções dos sistemas homogêneos em termos de conjuntos. Para isso, definimos o *conjunto solução* de $AX = 0$ por

$$S_A = \{v \in \mathbb{R}^n \mid Av = 0\},$$

em que, como sempre, estamos identificando a n -upla que descreve um vetor do \mathbb{R}^n com a respectiva matriz $n \times 1$. Observe que o critério a ser aplicado para determinar se um vetor v do \mathbb{R}^n pertence a S_A consiste em verificar que o produto Av é igual a zero. Com isto, as duas propriedades que encontramos anteriormente podem ser reformuladas como

- 1'. se $v_1, v_2 \in S_A$, então $v_1 + v_2 \in S_A$;
- 2'. se $v \in S_A$ e $\lambda \in \mathbb{R}$, então $\lambda v \in S_A$.

Os conjuntos que satisfazem estas duas propriedades são tão importantes que merecem um nome especial, são chamados de subespaços do \mathbb{R}^n .

3.2. Definição e mais exemplos. A definição oficial de subespaços que usaremos é a seguinte. Um subconjunto não-vazio U do \mathbb{R}^n é um *subespaço* se, quaisquer que sejam os vetores $v, v' \in U$ e o escalar λ , temos que

$$v + v' \in U \quad \text{e} \quad \lambda v \in U.$$

Note que um subespaço tem que conter o vetor zero, porque se v é um vetor qualquer de U e $\lambda = 0$, temos que $0 \cdot v = 0 \in U$. Isso faz com que, o menor de todos os subespaços do \mathbb{R}^n seja o conjunto cujo único elemento é o vetor zero. Por outro lado, o maior subespaço do \mathbb{R}^n é ele próprio.

Os sistemas homogêneos não são a única maneira de definirmos subespaços do \mathbb{R}^n . Também podemos fazê-lo a partir de geradores. Para isso, sejam w_1, \dots, w_m vetores do \mathbb{R}^n e considere o conjunto

$$W = \langle w_1, \dots, w_m \rangle$$

formado por todas as combinações lineares possíveis destes vetores. Portanto, o critério para saber se um vetor do \mathbb{R}^n pertence a W é determinar se podemos escrevê-lo como combinação linear de w_1, \dots, w_m . Assim, se v e v' estão em W , então existem números reais a_1, \dots, a_m e b_1, \dots, b_m tais que

$$v = a_1 w_1 + \cdots + a_m w_m \quad \text{e} \quad v' = b_1 w_1 + \cdots + b_m w_m.$$

Somando estes vetores, obtemos

$$v + v' = (a_1 + b_1)w_1 + \cdots + (a_m + b_m)w_m;$$

o que nos permite concluir que $v + v' \in W$. Como, pela distributividade do produto por escalar,

$$\lambda v = (\lambda a_1)w_1 + \cdots + (\lambda a_m)w_m,$$

também temos que $\lambda v \in W$. Consequentemente, W é um subespaço do \mathbb{R}^n .

Se considerarmos o que acontece no plano, logo descobrimos que as duas construções que temos para produzir subespaços não nos levam muito longe. Afinal, os sistemas homogêneos do \mathbb{R}^2 cuja matriz está em forma escada limitam-se àqueles definidos por uma única equação não nula. Digamos que a equação seja

$$a_1x_1 + a_2x_2 = 0.$$

Mas, como vimos no capítulo 1, esta equação define a reta que tem $(a_2, -a_1)$ como vetor diretor. Reciprocamente, a reta gerada pelo vetor não nulo (b_1, b_2) é solução do sistema homogêneo cuja equação é

$$b_2x_1 - b_1x_2 = 0.$$

Assim, as duas construções nos dão apenas retas. Contudo, há muitos outros subconjuntos do \mathbb{R}^n que nos são familiares, entre eles circunferências, parábolas, elipses e hipérboles. Será que um desses não poderia ser um subespaço? Vejamos o que acontece quando confrontamos estes conjuntos com o fato de que, multiplicando um vetor de um subespaço por um escalar nunca saímos do subespaço. Isso significa que um subespaço contém todas as retas geradas a partir de seus vetores não nulos. Contudo, uma reta qualquer não pode intersectar nenhuma das curvas que mencionamos em mais de dois pontos, o que nos permite descartar todas elas como exemplos de subespaços sem a necessidade de nenhuma verificação adicional.

Na verdade, não há outra maneira de gerar novos subespaços, além das duas que já consideramos, que não passam de duas maneiras diferentes de definir os mesmos conjuntos. Mas estas afirmações só serão provadas no artigo 6.4.

3.3. Eliminando redundâncias. No artigo anterior vimos que é possível definir um subespaço a partir de um conjunto finito de vetores, seus geradores. Entretanto é importante enfatizar que um subespaço não tem um único conjunto de geradores: na verdade, há uma infinidade de conjuntos de geradores para um mesmo subespaço. Uma maneira óbvia de justificar esta afirmação consiste em observar que podemos repetir um mesmo gerador várias vezes sem que nenhum novo vetor seja gerado como consequência disto. Uma resposta mais inteligente e mais significativa é dada pela proposição abaixo.

PROPOSIÇÃO 3.1. *Se w_1, \dots, w_n são vetores do \mathbb{R}^n e $b \in \mathbb{R}$, então*

$$\langle w_1, w_2 + bw_1, \dots, w_m \rangle = \langle w_1, w_2, \dots, w_m \rangle.$$

DEMONSTRAÇÃO. Vou provar primeiro que

$$(98) \quad \langle w_1, w_2 + bw_1, \dots, w_m \rangle \subset \langle w_1, w_2, \dots, w_m \rangle.$$

Se $v \in \langle w_1, w_2 + bw_1, \dots, w_m \rangle$, então

$$v = a_1w_1 + a_2(w_2 + bw_1) + a_3w_3 + \cdots + a_mw_m,$$

em que $a_1, \dots, a_m \in \mathbb{R}$. Logo,

$$v = (a_1 + a_2 b)w_1 + a_2 w_2 + a_3 w_3 + \cdots + a_m w_m$$

Portanto, $v \in \langle w_1, w_2, \dots, w_m \rangle$. Para provar a inclusão oposta, renomeamos os vetores

$$u_1 = w_1, u_2 = w_2 + bw_1, u_3 = w_3, \dots, u_m = w_m,$$

de modo que

$$w_2 = u_2 - bu_1 = u_2 - bu_1.$$

Com isso, a inclusão desejada equivale a

$$\langle u_1, u_2 - bu_1, \dots, u_m \rangle \subset \langle u_1, u_2, \dots, u_m \rangle.$$

que pode ser provada exatamente como procedemos para mostrar (98). \square

Quando o conjunto de geradores do subespaço tem mais de um vetor, esta proposição nos permite construir uma infinidade de novos conjuntos de geradores sem apelar para nenhuma estratégia repetições de vetores ou outra estratégia igualmente óbvia. Mas a real importância desta proposição está em que ela nos permite remover de um conjunto de geradores aqueles que são redundantes.

COROLÁRIO 3.2. *Seja G um conjunto finito de vetores do \mathbb{R}^n . As linhas da matriz escada obtida aplicando eliminação a $A(G)$ geram o mesmo subespaço que G .*

DEMONSTRAÇÃO. Para entender como isto é feito, considere o subespaço W do \mathbb{R}^n gerado pelos vetores do conjunto $G = \{w_1, \dots, w_k\}$. Associaremos a G a matriz

$$(99) \quad A(G) = \begin{bmatrix} \text{---} & w_1 & \text{---} \\ \text{---} & w_2 & \text{---} \\ \vdots & \ddots & \vdots \\ \text{---} & w_k & \text{---} \end{bmatrix},$$

de tamanho $k \times n$, cujas linhas são os vetores de G . Aplicamos, então, o método de eliminação gaussiana à matriz $A(G)$. Digamos que, executando o primeiro passo da eliminação, obtemos

$$\begin{bmatrix} \text{---} & w_1 & \text{---} \\ \text{---} & w_2 - \lambda w_1 & \text{---} \\ \vdots & \ddots & \vdots \\ \text{---} & w_k & \text{---} \end{bmatrix}$$

em que $\lambda \in \mathbb{R}$. Mas, de acordo com a proposição 3.1,

$$\langle w_1, w_2, \dots, w_k \rangle = \langle w_1, w_2 + \lambda w_1, \dots, w_k \rangle.$$

Contudo, o processo de eliminação gaussiana consiste em uma sucessão de passos análogos a este, além de possíveis trocas de posição das linhas, que em nada afetam o fato de termos

um conjunto de geradores. Portanto, se u_1, \dots, u_k são as linhas não nulas da matriz escada U obtida ao final da eliminação, podemos afirmar que

$$\langle w_1, w_2, \dots, w_k \rangle = \langle u_1, u_2, \dots, u_m \rangle,$$

para algum inteiro positivo $m \leq k$. \square

A presença de redundância se manifesta na redução da quantidade de vetores no conjunto de geradores obtido ao final da eliminação. Por exemplo, seja W o subespaço do \mathbb{R}^4 gerado pelos vetores

$$(2, 1, 3, 4), (4, 5, 8, 9), (-16, -17, -31, -35), (12, 12, 23, 26) \quad \text{e} \quad (20, 22, 39, 44).$$

Dispondo estes vetores como linhas em uma matriz obtemos

$$\begin{bmatrix} 2 & 1 & 3 & 4 \\ 4 & 5 & 8 & 9 \\ -16 & -17 & -31 & -35 \\ 12 & 12 & 23 & 26 \\ 20 & 22 & 39 & 44 \end{bmatrix}$$

cuja forma escada é

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

Portanto, dois, dos cinco vetores originais, eram redundantes e W pode ser igualmente gerado por

$$(2, 1, 3, 4), (0, 3, 2, 1), (0, 0, -1, 0).$$

4. Bases e dimensão

Na seção anterior vimos que é possível eliminar redundâncias no conjunto de geradores de um subespaço, de modo a obter um novo conjunto de geradores, mais simples, para o mesmo espaço. Contudo, não analisamos se o conjunto de geradores assim obtido é o menor possível ou se ainda daria para dispensar alguns vetores. Antes de podermos responder a esta pergunta precisamos definir precisamente o que significa um conjunto de geradores ser “o menor possível” ou, para usar a terminologia usual, ser uma base de um subespaço.

4.1. Definição e exemplos. Um conjunto finito de geradores β de um subespaço U é *base* se nenhum de seus vetores pode ser escrito como combinação linear dos demais. Consequentemente, o subespaço gerado pelo conjunto obtido removendo-se um dos vetores de β não é mais igual a U .

Para $1 \leq i \leq n$, seja e_i o vetor do \mathbb{R}^n que tem todas as entradas nulas, exceto pela i -ésima, que é igual a um. Como

$$(a_1, \dots, a_n) = a_1e_1 + \cdots + a_ne_n,$$

os vetores do conjunto

$$\varepsilon = \{e_1, \dots, e_n\}$$

geram \mathbb{R}^n . Além disso, e_i não pode ser escrito como combinação linear dos demais vetores, porque todos eles têm i -ésima entrada nula. Portanto, de acordo com a definição do parágrafo anterior, ε é uma base do \mathbb{R}^n . Como no caso do plano, esta base é conhecida como *base canônica* do \mathbb{R}^n , ainda que fosse mais correto chamá-la de base tautológica, porque não há nenhum conceito concreto de vetor que serve de lastro para a construção das n -uplas de coordenadas.

A resposta à pergunta feita na introdução desta seção é que as linhas não nulas de uma matriz em forma escada realmente definem uma base do subespaço gerado pelas linhas da matriz. Porém, antes de provar isto, precisamos reformular a definição de base de maneira mais apropriada aos cálculos necessários para verificar que um dado conjunto de geradores é uma base de um subespaço.

Sejam, então, w_1, \dots, w_k geradores de um subespaço U do \mathbb{R}^n . Suponhamos, para começar, que

$$G = \{w_1, \dots, w_k\},$$

não é uma base de U . Neste caso um deles pode ser escrito como combinação linear dos demais. Reordenando os vetores, se necessário, podemos supor que este vetor é w_k . Logo, existem números reais a_1, \dots, a_{k-1} tais que

$$w_k = a_1w_1 + \cdots + a_{k-1}w_{k-1}.$$

Passando o w_k para o lado direito da equação, obtemos

$$a_1w_1 + \cdots + a_{k-1}w_{k-1} - w_k = 0.$$

Logo, se o conjunto G não for base, existe uma combinação linear dos seus elementos que é igual a zero, *sem que todos os coeficientes da combinação linear sejam nulos*. De fato, no argumento acima, o coeficiente de w_k é um e, portanto, não nulo. Dizemos que um conjunto de vetores com esta propriedade é *linearmente dependente*. Como seria de esperar, um conjunto que não é linearmente dependente é chamado de *linearmente independente*. Com isso podemos enunciar a caracterização de base de que precisamos.

LEMA 4.1. *Um conjunto finito de geradores de um espaço vetorial é uma base se, e somente se, for linearmente independente.*

DEMONSTRAÇÃO. O argumento acima mostra que, se um conjunto de geradores não é uma base, então tem que ser linearmente dependente. Para obter o resultado do lema falta apenas provar a recíproca. Suponhamos, então, que $G = \{w_1, \dots, w_k\}$ seja um conjunto linearmente dependente de vetores que geram um subespaço U . Neste caso, existem números reais a_1, \dots, a_k tais que

$$a_1 w_1 + \cdots + a_k w_k = 0$$

é uma combinação linear dos vetores de G com algum coeficiente não nulo. Reordenando os vetores de G , podemos supor que $a_k \neq 0$. Logo,

$$w_k = -\frac{a_1}{a_k} w_1 - \cdots - \frac{a_{k-1}}{a_k} w_{k-1},$$

mostrando, assim, que G não é base. \square

Com isso estamos preparados para provar que o conjunto de geradores construído a partir do processo de eliminação gaussiana é, de fato, uma base do subespaço gerado pelas linhas da matriz.

PROPOSIÇÃO 4.2. *Seja U um subespaço do \mathbb{R}^n e G um conjunto finito de geradores de U que não contém o vetor nulo. Se a matriz $A(G)$ está em forma escada, então G é uma base de U .*

DEMONSTRAÇÃO. Pelo lema 4.1, basta provar que

$$G = \{w_1, \dots, w_m\}.$$

é linearmente independente. Suponhamos, por contradição, que

$$(100) \quad b_1 w_1 + \cdots + b_m w_m = 0$$

e que ℓ é o menor inteiro positivo para o qual $b_\ell \neq 0$. Seja

$$w_\ell = (\underbrace{0, \dots, 0}_{k-1 \text{ posições}}, a_k, a_{k+1}, \dots, a_n)$$

com $a_k \neq 0$. Como a matriz $A(G)$ está em forma escada, isto significa que as primeiras k entradas de

$$w_{\ell+1}, \dots, w_m$$

são todas nulas. Logo, a k -ésima entrada na expressão do lado esquerdo de (100) é igual a $b_\ell a_k = 0$, contradizendo nossa hipótese de que b_ℓ e a_k são ambos diferentes de zero, o que completa a demonstração da proposição. \square

Esta proposição sugere que todos os subespaços do \mathbb{R}^n admitem uma base. Apesar do fato em si ser verdadeiro, não podemos deduzi-lo do que fizemos até aqui. A razão, é que tudo o que fizemos pressupõe que tenhamos à mão um conjunto finito de geradores aos quais aplicar a eliminação; mas nada do que fizemos indica que isto seja sempre verdadeiro.

Fecharemos esta brecha em nossa argumentação no teorema que abre o artigo 4.3. Antes, porém, precisamos provar o resultado chave do qual depende aquele teorema.

4.2. Lema da troca. Todo este artigo será dedicado a um único lema e à sua demonstração. O resultado é um dos mais importantes do livro; mas a demonstração, baseada no *argumento da troca* de Ernest Steinitz é ainda mais importante, porque se aplica a outras situações e deveria estar na caixa de ferramentas de qualquer pessoa interessada em matemática. Por isso optei por analisar detalhadamente esta demonstração, em vez de usar um caminho mais direto. Deste lema depende tudo o mais que faremos sobre bases.

LEMA DA TROCA. *Seja U um subespaço do \mathbb{R}^n . Se U admite uma base com m elementos, então um conjunto finito e linearmente independente contido em U tem, no máximo, m elementos.*

DEMONSTRAÇÃO. Sejam

$$\beta = \{u_1, \dots, u_m\}$$

uma base de U e

$$v_1, \dots, v_r$$

vetores linearmente independentes em U . Nossa objetivo é mostrar que $r \leq m$. A estratégia que vamos adotar para atingir esse objetivo consiste em trocar os elementos de β por v 's de forma que o conjunto resultante nunca deixe de ser uma base de U .

Esboço da demonstração. Suponhamos, por contradição, que $r > m$. Admitindo que a estratégia acima seja viável, partimos de

$$\beta_0 = \{u_1, \dots, u_m\} \quad \text{e} \quad \alpha_0 = \{v_1, \dots, v_r\},$$

e reordenando os u 's, se necessário, substituímos u_1 por v_1 em β_0 e removeremos v_1 de α_0 , obtendo uma nova base

$$\beta_1 = \{v_1, u_2, \dots, u_m\}$$

para U , ao passo que

$$\alpha_1 = \{v_2, \dots, v_r\},$$

Repetindo o processo trocamos, desta vez, u_2 por v_2 , de modo que a base de U será

$$\beta_2 = \{v_1, v_2, u_3, \dots, u_m\}$$

e teremos

$$\alpha_2 = \{v_3, \dots, v_r\}.$$

Continuando desta maneira, ao cabo de m etapas descobriremos que

$$\beta_m = \{v_1, v_2, \dots, v_m\}$$

é uma base de U . Além disso, como estamos supondo que $r > m$,

$$\alpha_m = \{v_{m+1}, \dots, v_r\} \neq \emptyset.$$

Contudo, se β_m é base de U , então $v_{m+1} \in U$ pode ser escrito como combinação linear dos vetores em v_1, v_2, \dots, v_m . Mas isto significa que

$$v_1, v_2, \dots, v_m, v_{m+1}$$

são vetores linearmente dependentes, contradizendo uma das hipóteses do lema. Logo, $r \leq m$, como devíamos mostrar.

A receita recursiva. Agora que entendemos como a demonstração funciona em linhas gerais, podemos descrever qual será a receita usada para realizar as trocas de vetores de modo que

$$\beta_1, \beta_2, \dots, \beta_m$$

continuem sendo bases de U . Para isto, digamos que tivemos sucesso em obter uma base

$$\beta_{k-1} = \{v_1, \dots, v_{k-1}, u_k, \dots, u_m\}$$

para U , com

$$\alpha_{k-1} = \{v_k, \dots, v_r\}.$$

A receita recursiva que usaremos para construir β_k consiste de três passos:

Passo 1: como β_{k-1} é base de U , existem números reais a_1, \dots, a_m tais que

$$(101) \quad v_k = a_1 v_1 + \dots + a_{k-1} v_{k-1} + a_k u_k + \dots + a_m u_m$$

Passo 2: reordene u_k, \dots, u_m , se necessário, de modo que $a_k \neq 0$;

Passo 3: troque u_k por v_k em β_{k-1} e remova v_k de α_{k-1} , obtendo

$$\beta_k = \{v_1, \dots, v_k, u_{k+1}, \dots, u_m\} \quad \text{e} \quad \alpha_k = \{v_{k+1}, \dots, v_r\}.$$

Note que, no passo 2, só reordenamos u_k, \dots, u_m ; deixando v_1, \dots, v_{k-1} de fora da reordenação. Portanto, para que este passo possa ser executado é necessário que haja pelo menos um número não nulo entre os coeficientes a_k, \dots, a_m . Isto ocorre, porque se todos estes coeficientes fossem todos nulos, a equação (101) nos daria v_k como combinação linear de v_1, \dots, v_{k-1} , contradizendo a hipótese de que os v 's são linearmente independentes.

Demonstração de que a receita recursiva produz bases. Para encerrar, precisamos provar que o conjunto β_k construído usando a receita recursiva é uma base de U . Em primeiro lugar, temos de (101) que

$$u_k = -\frac{a_1}{a_k} v_1 - \dots - \frac{a_{k-1}}{a_k} v_{k-1} + \frac{1}{a_k} v_k - \frac{a_{k+1}}{a_k} u_{k+1} - \dots - \frac{a_m}{a_k} u_m;$$

onde como o conjunto

$$u_k \in \langle v_1, \dots, v_k, u_{k+1}, \dots, u_m \rangle.$$

Logo,

$$U = \langle v_1, \dots, v_{k-1}, u_k, u_{k+1}, \dots, u_m \rangle \subset \langle v_1, \dots, v_k, u_{k+1}, \dots, u_m \rangle \subset U,$$

de modo que β_k é um conjunto de geradores de U . Finalmente, suponhamos que

$$b_1v_1 + \cdots + b_kv_k + b_{k+1}u_{k+1} + \cdots + b_mu_m = 0.$$

Substituindo a fórmula para v_k em (101) na expressão acima e agrupando as parcelas que multiplicam um mesmo vetor, obtemos

$$\left(\sum_{i=1}^k (b_i - b_k a_i) v_i \right) + b_k a_k u_k + \left(\sum_{j=k+1}^m (b_j - b_k a_j) u_j \right) = 0.$$

Logo,

$$b_k a_k = 0 \quad \text{e} \quad (b_i - b_k a_i) = 0 \quad \text{para } i = 1, \dots, k-1, k+1, \dots, m.$$

Porém, como, pela receita recursiva, $a_k \neq 0$, então $b_k a_k = 0$ nos permite concluir que $b_k = 0$. Substituindo isto nos demais coeficientes, resta

$$b_i = 0 \quad \text{para } i = 1, \dots, k-1, k+1, \dots, m.$$

Portanto, os vetores em β_k são linearmente independentes, completando, assim, a demonstração de que β_k é uma base de U . \square

Para tornar o procedimento mais concreto, vejamos como o lema da troca seria aplicado aos subconjuntos

$$\begin{aligned} \beta_0 &= \{ \underbrace{(0, 0, 1)}_{u_1}, \underbrace{(1, 1, 0)}_{u_2}, \underbrace{(1, 0, 1)}_{u_3} \} \\ \alpha_0 &= \{ \underbrace{(1, 1, 1)}_{v_1}, \underbrace{(0, 1, 1)}_{v_2}, \underbrace{(0, 0, 1)}_{v_3} \} \end{aligned}$$

do \mathbb{R}^3 . Como

$$\underbrace{(1, 1, 1)}_{v_1} = \underbrace{(1, 1, 0)}_{u_2} + \underbrace{(0, 0, 1)}_{u_1},$$

posso trocar $(1, 1, 0)$ por $(1, 1, 1)$. Reordenando os vetores de β_0 para pôr $(1, 1, 0)$ na primeira posição, ficamos com

$$\begin{aligned} \beta_0 &= \{ \underbrace{(1, 1, 0)}_{u_1}, \underbrace{(0, 0, 1)}_{u_2}, \underbrace{(1, 0, 1)}_{u_3} \} \\ \alpha_0 &= \{ \underbrace{(1, 1, 1)}_{v_1}, \underbrace{(0, 1, 1)}_{v_2}, \underbrace{(0, 0, 1)}_{v_3} \}. \end{aligned}$$

Fazendo, agora, a troca

$$\begin{aligned} \beta_1 &= \{ \underbrace{(1, 1, 1)}_{v_1}, \underbrace{(0, 0, 1)}_{u_2}, \underbrace{(1, 0, 1)}_{u_3} \} \\ \alpha_1 &= \{ \underbrace{(0, 1, 1)}_{v_2}, \underbrace{(0, 0, 1)}_{v_3} \}. \end{aligned}$$

Em seguida, vejo que

$$\underbrace{(0, 1, 1)}_{v_2} = \underbrace{(1, 1, 1)}_{v_1} + \underbrace{(0, 0, 1)}_{u_2} - \underbrace{(1, 0, 1)}_{u_3}$$

Note que, pelo passo recursivo, só posso trocar v_2 por u_2 ou u_3 . Trocando por u_2 , obtemos

$$\begin{aligned}\beta_2 &= \left\{ \underbrace{(1, 1, 1)}_{v_1}, \underbrace{(0, 1, 1)}_{v_2}, \underbrace{(1, 0, 1)}_{u_3} \right\} \\ \alpha_2 &= \left\{ \underbrace{(0, 0, 1)}_{v_3} \right\}.\end{aligned}$$

Finalmente,

$$\underbrace{(0, 0, 1)}_{v_3} = - \underbrace{(1, 1, 1)}_{v_1} + \underbrace{(0, 1, 1)}_{v_2} + \underbrace{(1, 0, 1)}_{u_3}$$

e só posso trocar $(1, 0, 1)$ por $(0, 0, 1)$; donde

$$\begin{aligned}\beta_3 &= \left\{ \underbrace{(1, 1, 1)}_{v_1}, \underbrace{(0, 1, 1)}_{v_2}, \underbrace{(0, 0, 1)}_{v_3} \right\} \\ \alpha_3 &= \emptyset,\end{aligned}$$

completando assim o procedimento de troca.

4.3. Dimensão. O teorema abaixo contém algumas mais importantes propriedades das bases.

TEOREMA 4.3. *Se U é um subespaço do \mathbb{R}^n , então*

- (a) *U admite uma base;*
- (b) *quaisquer duas bases de U têm a mesma quantidade de elementos;*
- (c) *qualquer subconjunto linearmente independente de U pode ser completado para uma base de U .*

DEMONSTRAÇÃO. Se U contém apenas o vetor nulo, não há nada a fazer, de modo que podemos supor que $U \neq \{0\}$. Como $U \subset \mathbb{R}^n$ e \mathbb{R}^n tem uma base com n elementos, sabemos, pelo lema da troca, que U não pode conter um conjunto linearmente independente com mais de n elementos. Seja β um subconjunto do \mathbb{R}^n com a maior quantidade possível de elementos linearmente independentes. Neste caso, qualquer vetor $v \in U$ é combinação linear dos vetores de β ; caso contrário, $\beta \cup \{v\}$ seria um conjunto linearmente independente maior que β , o que não é possível pela maximalidade de β . Portanto, β é base de U e provamos (a).

Para provar (b), digamos que β_1 e β_2 são duas bases de um mesmo subespaço U do \mathbb{R}^n . Como β_1 é base de U e β_2 é um conjunto linearmente independente contido em U , temos, pelo lema da troca, que $\#\beta_2 \leq \#\beta_1$. Contudo, também é verdade que β_2 é base de U e que

β_1 é um conjunto linearmente independente em U , de modo que o Lema da Troca também nos permite concluir que $\#\beta_1 \leq \#\beta_2$. Portanto, $\#\beta_1 = \#\beta_2$ e provamos também (b).

Seja α um conjunto linearmente independente contido em U . A demonstração de (c) é baseada no fato de que, ou α gera U ou existe um vetor em U que não é combinação linear dos vetores de α . Neste último caso, este vetor pode ser acrescentado a α sem que o novo conjunto deixe de ser linearmente independente. Contudo, pelo lema da troca, este último passo não pode ser aplicado mais do que n vezes. Portanto, em algum momento o conjunto α terá se tornado uma base de U . \square

Pelo teorema, a quantidade de elementos na base de um subespaço $U \neq \{0\}$ do \mathbb{R}^n não depende da base escolhida, mas apenas do subespaço U . Isto nos permite associar a U o número inteiro positivo correspondente à quantidade de elementos de qualquer de suas bases. Este número é chamado de *dimensão* de U e será denotado por $\dim(U)$. Como o subespaço $\{0\}$ não se adequa a esta definição, convencionamos que tem dimensão igual a zero. A seguinte propriedade da dimensão é consequência direta do item (c) do teorema.

COROLÁRIO 4.4. *Se $U \subsetneq W$ são subespaços do \mathbb{R}^n , então $\dim(U) < \dim(W)$.*

Quando um subespaço é definido como conjunto solução de um sistema homogêneo, sua dimensão pode ser calculada diretamente a partir da forma escada da matriz do sistema. Para entender o que está em jogo, começaremos analisando o conjunto $S \subset \mathbb{R}^6$ das soluções do sistema homogêneo.

$$(102) \quad \begin{cases} x_1 + x_2 - x_4 - 2x_6 = 0 \\ x_2 - 3x_5 = 0 \\ x_4 + 7x_6 = 0. \end{cases}$$

Como a matriz deste sistema já está em forma escada, podemos resolvê-lo por substituição. Das duas últimas equações obtemos

$$x_2 = 3x_5 \quad \text{e} \quad x_4 = -7x_6$$

Substituindo isto na primeira equação, verificamos que

$$x_1 = -3x_5 - 5x_6.$$

Portanto, $v = (x_1, x_2, x_3, x_4, x_5)$ é solução de (102) se, e somente se,

$$(103) \quad v = (-3x_5 - 5x_6, 3x_5, x_3, -7x_6, x_5, x_6),$$

que, por sua vez, é igual à combinação linear

$$(104) \quad v = x_3(0, 0, 1, 0, 0, 0) + x_5(-3, 3, 0, 0, 1, 0) + x_6(-5, 0, 0, -7, 0, 1).$$

Logo, S pode ser gerado pelos vetores do conjunto

$$G = \{(0, 0, 1, 0, 0, 0), (-3, 3, 0, 0, 1, 0), (-5, 0, 0, -7, 0, 1)\}$$

Para saber se este conjunto é uma base precisamos descobrir se estes vetores são linearmente independentes. Para isto basta supor que $v = 0$ em (104) e determinar os valores que x_3, x_5 e x_6 podem tomar para que isto aconteça. Mas isto equivale a igualar v a zero em (103), donde obtemos

$$x_3 = x_5 = x_6 = 0.$$

Como os vetores de G geram S e são linearmente independentes, G é uma base de S . Logo,

$$\dim(S) = \#G = 3$$

é igual à diferença entre $\dim(\mathbb{R}^6) = 6$ e a quantidade de equações do sistema.

PROPOSIÇÃO 4.5. *Seja A uma matriz escada $m \times n$. A dimensão do conjunto solução do sistema $AX = 0$ no \mathbb{R}^n é igual a n menos a quantidade de linhas não nulas de A .*

DEMONSTRAÇÃO. Seja

$$U = \{v \in \mathbb{R}^n \mid Av = 0\}.$$

Como o conjunto solução de um sistema linear e de sua redução à forma escada são equivalentes, podemos supor que A está em forma escada com ℓ linhas não nulas. Para $1 \leq j \leq \ell$, denotaremos por m_j a variável que lidera a j -ésima linha do sistema $AX = 0$. Portanto, resolvendo o sistema por substituição, obtemos uma solução paramétrica da forma

$$(105) \quad x_{m_i} = \sum_{j \notin I} \alpha_{i,j} x_j$$

em que $I = \{m_1, \dots, m_\ell\}$. Note que estamos usando $j \notin I$ como abreviação de

$$j \in \{1, \dots, \ell\} \setminus I.$$

Substituindo as expressões em (105) no vetor $v = (x_1, \dots, x_n)$, podemos escrever

$$(106) \quad v = \sum_{j \in I} x_j u_j$$

em que

$$u_j = (b_{1,j}, \dots, b_{n,j}) \text{ e } b_{k,j} = \begin{cases} \alpha_{i,j} & \text{quando } k = m_i \\ 1 & \text{quando } k = j \\ 0 & \text{para qualquer outro valor de } k. \end{cases}$$

Isto mostra que o conjunto

$$G = \{u_j \mid j \in I\}$$

gera U . Finalmente, para cada $j \notin I$, existe um único vetor de G com a j -ésima entrada não nula, o que nos permite concluir que fazendo $v = 0$ em (106) obteremos

$$x_j = 0 \text{ para todo } j \notin I;$$

o que nos permite concluir que os vetores de G são linearmente independentes. \square

Encerraremos o capítulo com duas consequências importantes desta proposição. A primeira diz respeito aos hiperplanos, que foram definidos no artigo 2.2, como o conjunto solução de uma única equação não nula no \mathbb{R}^n .

COROLÁRIO 4.6. *Um hiperplano do \mathbb{R}^n tem dimensão igual a $n - 1$.*

Como a demonstração deste primeiro corolário é muito simples, deixaremos a tarefa de prová-lo por sua conta. A segunda consequência da proposição 4.5 está relacionada à quantidade de linhas não nulas na forma escada de uma dada matriz.

COROLÁRIO 4.7. *Todas as formas escada de uma dada matriz têm a mesma quantidade de linhas não nulas.*

DEMONSTRAÇÃO. Sejam E_1 e E_2 duas formas escada de uma matriz A com n colunas. Como vimos no capítulo 4, o conjunto solução de um sistema homogêneo e do sistema correspondente à forma escada de sua matriz são iguais. Logo,

$$\dim(S_A) = \dim(S_{E_2}) = \dim(S_{E_1}).$$

Denotando por r_1 e r_2 a quantidade de linhas não nulas em E_1 e E_2 , temos pela proposição 4.5, que

$$r_1 = n - \dim(S_{E_1}) = n - \dim(S_{E_2}) = r_2,$$

como queríamos mostrar. □

Portanto, de acordo com este último corolário, a quantidade de linhas não nulas na forma escada de uma matriz A depende apenas da matriz e não da forma escada escolhida. Esta quantidade é conhecida como o *posto* de A e denotada por $\text{posto}(A)$. Com isto podemos formular a proposição compactamente pela igualdade

$$\dim(S_A) = n - \text{posto}(A),$$

em que n é a quantidade de colunas de A .

5. Bases ortonormais

Tendo introduzido bases, veremos nesta seção como generalizar o conceito de base ortonormal do plano para o \mathbb{R}^n e como proceder para calcular uma base ortonormal de um dado subespaço.

5.1. Definição e exemplos. Como no caso do plano, diremos que o conjunto

$$\beta = \{u_1, \dots, u_m\}$$

é *base ortonormal* de um subespaço U do \mathbb{R}^n se estes vetores geram U e satisfazem

$$(107) \quad \langle u_i | u_j \rangle = \begin{cases} 1 & \text{se } i = j, \\ 0 & \text{se } i \neq j. \end{cases}$$

Antes de prosseguir há um detalhe importante que precisa ser esclarecido. Para que possamos nos referir a β como base é necessário, segundo o critério do lema 4.1, que seus vetores sejam linearmente independentes. No caso das bases ortonormais, a independência linear dos vetores é consequência de (107). Para mostrar isto, seja

$$b_1 u_1 + \cdots + b_m u_m = 0$$

uma combinação linear nula dos vetores de β . Calculando o produto interno dos dois lados desta expressão por u_k , obtemos, pelas propriedades do produto interno, que

$$b_1 \langle u_1 | u_k \rangle + \cdots + b_m \langle u_m | u_k \rangle = 0.$$

Logo, por (107), $b_k = 0$. Como isto vale para todo $1 \leq k \leq m$, podemos concluir que os vetores de β são linearmente independentes. Tendo tirado isto do caminho, vejamos alguns exemplos.

Em primeiro lugar, qualquer que seja o inteiro $n \geq 1$, a base canônica é uma base ortonormal do \mathbb{R}^n . Para nosso segundo exemplo calcularemos uma base ortonormal para o plano H do \mathbb{R}^3 cuja equação é

$$x_1 - x_2 + x_3 = 0.$$

Os vetores de H são da forma

$$(108) \quad (x_2 - x_3, x_2, x_3) = x_2(1, 1, 0) + x_3(-1, 0, 1).$$

Logo, os vetores

$$(1, 1, 0) \quad \text{e} \quad (-1, 0, 1),$$

formam uma base de H . Contudo, esta não é uma base ortonormal, porque o produto interno destes vetores é igual a um. A estratégia que adotaremos para construir uma base ortonormal para H consiste em escolher um vetor não nulo qualquer de H como ponto de partida e, então, determinar os vetores de H que são ortogonais ao vetor escolhido. Digamos que escolhemos

$$u_1 = \frac{1}{\sqrt{2}}(1, 1, 0).$$

Usando (108), verificamos que um vetor qualquer $(x_2 - x_3, x_2, x_3)$ em H é ortogonal a u_1 se

$$0 = \langle (x_2 - x_3, x_2, x_3) | (1, 1, 0) \rangle = 2x_2 - x_3.$$

Logo, para obter os vetores de H ortogonais a u_1 , basta substituir $x_3 = 2x_2$ em (108), o que nos dá

$$(-x_2, x_2, 2x_2).$$

Por exemplo,

$$u_2 = \frac{1}{\sqrt{6}}(-1, 1, 2)$$

é unitário e ortogonal a u_1 . Como H tem dimensão dois, podemos concluir que $\{u_1, u_2\}$ é uma base ortonormal de H . Por outro lado, pela definição de hiperplano, sabemos que o vetor unitário

$$u_3 = \frac{1}{\sqrt{3}}(1, -1, 1)$$

é ortogonal a todos os vetores de H . Assim, o que fizemos também nos mostra que $\{u_1, u_2, u_3\}$ é uma base ortonormal do \mathbb{R}^3 diferente da base canônica.

Uma estratégia análoga pode ser usada para calcular uma base ortonormal de qualquer subespaço de dimensão dois do \mathbb{R}^n . Contudo, se o subespaço tiver dimensão maior que dois, será necessário repetir este procedimento várias vezes até encontrar vetores suficientes para termos uma base do subespaço. Entretanto, em vez de darmos outro exemplo da mesma natureza, mas com dimensão maior, descreveremos um algoritmo eficiente que produz uma base ortonormal, a partir de qualquer conjunto de geradores de um subespaço.

5.2. Vetor ortogonal a um subespaço. Embora nosso objetivo seja descrever um algoritmo capaz de construir uma base ortonormal a partir de geradores de um subespaço, começaremos investigando um problema mais simples. Suponhamos que S seja um subespaço do \mathbb{R}^n do qual já conhecemos uma base ortonormal

$$\beta = \{u_1, \dots, u_m\}$$

e que $v \neq 0$ seja um vetor do \mathbb{R}^n que *não* pertence a S . Queremos descobrir como construir, a partir de v e de β , um vetor w ortogonal a S ; que é uma maneira sucinta de dizer que w deve ser ortogonal a *todos* os elementos de S . Faremos isto, subtraindo de v um vetor de S ; mais precisamente, determinaremos números reais a_1, \dots, a_m de modo que

$$w = v - (a_1 u_1 + \dots + a_m u_m)$$

seja ortogonal a todo vetor de S . Como os vetores de S são combinações lineares dos vetores de u_1, \dots, u_m , basta que w seja ortogonal aos vetores de β para que seja ortogonal a todos os vetores de S . Mas, calculando o produto interno entre w e $u_i \in \beta$, obtemos

$$\langle u_i | w \rangle = \langle u_i | v \rangle - a_1 \langle u_i | u_1 \rangle - \dots - a_m \langle u_i | u_m \rangle.$$

Levando em conta que β é uma base ortonormal, a equação anterior equivale a

$$\langle u_i | w \rangle = \langle u_i | v \rangle - a_i.$$

Portanto, w é ortogonal a u_i se, e somente se, $a_i = \langle u_i | v \rangle$, que nos dá o seguinte resultado.

PROPOSIÇÃO 5.1. *Se $\beta = \{u_1, \dots, u_m\}$ é uma base ortonormal de um subespaço S do \mathbb{R}^n e $v \notin S$, então*

$$v - \langle u_1 | v \rangle u_1 - \dots - \langle u_m | v \rangle u_m$$

é um vetor ortogonal a S .

Seja, por exemplo, S o subespaço do \mathbb{R}^5 gerado pelos vetores

$$u_1 = \frac{1}{2}(1, 1, 1, 1, 0) \quad \text{e} \quad u_2 = \frac{1}{2}(1, -1, 1, -1, 0),$$

que são ortogonais e unitários. É claro que $v = (6, 2, 4, 2, 1)$ não pertence a S , porque todos os vetores de S têm quinta coordenada igual a zero, o que não ocorre com v . Como

$$\langle v|u_1 \rangle = 7 \quad \text{e} \quad \langle v|u_2 \rangle = 3,$$

concluímos, pela proposição 5.1, que

$$v - 7u_1 - 3u_2 = (1, 0, -1, 0, 1)$$

é perpendicular a S .

5.3. O algoritmo de Gram-Schmidt. Seja S um subespaço do \mathbb{R}^n do qual conhecemos a base

$$\beta = \{v_1, \dots, v_m\}.$$

Construiremos uma base ortonormal β^* para S , um vetor de cada vez. Seja β_i^* a parte da base ortonormal construída até a i -ésima iteração do algoritmo. Para dar início à construção recursiva, tomamos

$$\beta_1^* = \{u_1\} \quad \text{em que} \quad u_1 = \frac{v_1}{\|v_1\|}.$$

A passagem de β_i^* a β_{i+1}^* é feita com a ajuda da proposição 5.1. Suponhamos que

$$\beta_i^* = \{u_1, \dots, u_i\}$$

seja uma base ortonormal do subespaço

$$S_i = \langle v_1, \dots, v_i \rangle.$$

Definindo u_{i+1} como a normalização do vetor

$$v_{i+1} - \sum_{j=1}^i \langle u_j | v_{i+1} \rangle u_j,$$

temos, pela proposição 5.1, que u_{i+1} é ortogonal aos vetores de S_i . Além disso, como u_{i+1} é combinação linear de u_1, \dots, u_i e v_{i+1} , segue-se que

$$S_{i+1} = \langle u_1, \dots, u_i, u_{i+1} \rangle = \langle u_1, \dots, u_i, v_{i+1} \rangle = \langle v_1, \dots, v_i, v_{i+1} \rangle,$$

em que a terceira igualdade é consequência da hipótese de que β_i^* é base de S_i . Em particular,

$$\beta_{i+1}^* = \{u_1, \dots, u_i, u_{i+1}\}$$

é uma base ortonormal do subespaço

$$S_{i+1} = \langle v_1, \dots, v_i, v_{i+1} \rangle,$$

para todo $1 \leq i \leq m-1$. Este procedimento é conhecido como *método de Gram-Schmidt* e podemos sistematizá-lo como abaixo.

ALGORITMO 5.2 (Método de Gram-Schmidt). *Dados vetores linearmente independentes $v_1, \dots, v_m \in \mathbb{R}^n$, o algoritmo retorna uma base ortonormal do subespaço $\langle v_1, \dots, v_m \rangle$.*

*inicialize β como a lista vazia;
atribua a u_1 a normalização de v_1 ;
para $i = 2, \dots, m$ faça:
 calcule $w_i = v_i - \langle v_i | u_1 \rangle u_1 - \dots - \langle v_i | u_{i-1} \rangle u_{i-1}$;
 atribua a u_i a normalização de w_i ;
 acrescente u_i ao final de β ;
retorne β .*

Considere, por exemplo, o hiperplano H do \mathbb{R}^4 cuja equação é

$$x - y - 2z - w = 0.$$

Os vetores de H são da forma

$$(x, y, z, w) = (y + 2z + w, y, z, w) = y(1, 1, 0, 0) + z(2, 0, 1, 0) + w(1, 0, 0, 1)$$

Portanto,

$$\beta = \{(1, 1, 0, 0), (2, 0, 1, 0), (1, 0, 0, 1)\}$$

é base de H . Usaremos β como ponto de partida para encontrar uma base ortonormal de H , através do algoritmo de Gram-Schmidt. A inicialização requer que façamos

$$\beta_1^* = \left\{ \frac{1}{\sqrt{2}}(1, 1, 0, 0) \right\},$$

que é a normalização do primeiro vetor de β . Para achar β_2^* , calculamos o produto interno

$$\left\langle (2, 0, 1, 0), \frac{1}{\sqrt{2}}(1, 1, 0, 0) \right\rangle = \frac{2}{\sqrt{2}} = \sqrt{2};$$

de modo que o segundo vetor de β_2^* será a normalização de

$$(2, 0, 1, 0) - \frac{\sqrt{2}}{\sqrt{2}}(1, 1, 0, 0) = (1, -1, 1, 0);$$

que é igual a

$$\frac{1}{\sqrt{3}}(1, -1, 1, 0).$$

Portanto,

$$\beta_2^* = \left\{ \frac{1}{\sqrt{2}}(1, 1, 0, 0), \frac{1}{\sqrt{3}}(1, -1, 1, 0) \right\},$$

Já o último vetor da base ortonormal é obtido normalizando

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

o que nos dá

$$\frac{1}{6}(1, -1, -2, 6).$$

Portanto, a base ortonormal de H calculada a partir de β pelo algoritmo de Gram-Schmidt é

$$\beta_3^* = \left\{ \frac{1}{\sqrt{2}}(1, 1, 0, 0), \frac{1}{\sqrt{3}}(1, -1, 1, 0), \frac{1}{\sqrt{42}}(1, -1, -2, 6) \right\}.$$

5.4. Ângulos entre vetores. Neste artigo definiremos o ângulo entre vetores do \mathbb{R}^n e obteremos uma fórmula que nos permite calculá-lo a partir do produto interno do \mathbb{R}^n , tal qual fizemos no caso do plano.

Sejam, então, $v \neq 0$ e $v' \neq 0$ vetores não colineares do \mathbb{R}^n . Como estamos supondo que estes vetores não são colineares, eles geram um plano do \mathbb{R}^n que chamaremos de U . O ângulo entre v e v' pode, então, ser definido como o menor ângulo entre estes vetores *medido sobre o plano U* . Reduzimos, assim, o problema de medir o ângulo entre vetores do \mathbb{R}^n ao problema de medir o ângulo entre vetores do plano, que aprendemos como fazer no ensino fundamental. Não há nada surpreendente nisto, uma vez que vetores desenhados em uma folha de papel são vetores do espaço tridimensional que habitamos, mas o ângulo entre eles é medido sobre a folha plana do papel.

Embora seja bastante intuitiva, esta maneira de medir o ângulo entre vetores do \mathbb{R}^n não é fácil de usar, por isso buscarmos uma fórmula mais adequada aos cálculos que precisaremos fazer no próximo capítulo. Continuando com a notação do parágrafo anterior, começaremos completando $\{v, v'\}$ para uma base do \mathbb{R}^n e aplicando a esta base o algoritmo de Gram-Schmidt. Obtemos, assim, uma base ortonormal β do \mathbb{R}^n cujos dois primeiros vetores, que chamaremos de u_1 e u_2 , formam uma base ortonormal do plano U . Como v e v' pertencem a U ,

$$v = a_1 u_1 + a_2 u_2 \quad \text{e} \quad v' = a'_1 u_1 + a'_2 u_2,$$

de modo que

$$\langle v | v' \rangle = \langle a_1 u_1 + a_2 u_2 | a'_1 u_1 + a'_2 u_2 \rangle.$$

Antes de continuarmos, note que o produto interno que estamos considerando é aquele definido no \mathbb{R}^n , que é o espaço ao qual pertencem v e v' . Mas, pela distributividade do produto interno,

$$\langle v | v' \rangle = a_1 a'_1 \langle u_1 | u_1 \rangle + a_1 a'_2 \langle u_1 | u_2 \rangle + a_2 a'_1 \langle u_2 | u_1 \rangle + a_2 a'_2 \langle u_2 | u_2 \rangle.$$

Levando em conta que β é uma base ortonormal do \mathbb{R}^n , esta última equação nos dá

$$(109) \quad \langle v | v' \rangle = a_1 a'_1 + a_2 a'_2.$$

Assim, o produto interno entre v e v' , como vetores do \mathbb{R}^n , é igual ao produto interno que teríamos obtido se tivéssemos considerado estes vetores meramente como vetores do plano

cujas coordenadas fossem, respectivamente, (a_1, a_2) e (a'_1, a'_2) . Em particular,

$$(110) \quad \|v\|^2 = \langle v | v \rangle = a_1^2 + a_2^2$$

e uma fórmula análoga vale para v' .

Porém, sabemos pelo artigo 1.3 do capítulo 1 que se θ é o menor ângulo entre (a_1, a_2) e (a'_1, a'_2) então

$$\cos(\theta) = \frac{a_1 a'_1 + a_2 a'_2}{\sqrt{a_1^2 + a_2^2} \sqrt{a'_1^2 + a'_2^2}}.$$

Portanto, segue-se de (109) e (110), que

$$\cos(\theta) = \frac{\langle v | v' \rangle}{\|v\| \|v'\|}.$$

Mostramos, assim, que o menor ângulo θ , medido entre v e v' considerados apenas como vetores no plano U , pode ser calculado utilizando apenas o produto interno no \mathbb{R}^n e a norma destes vetores, sem nenhuma referência ao plano U . Em particular, como $|\cos(\theta)| \leq 1$, podemos concluir que

$$\frac{|\langle v | v' \rangle|}{\|v\| \|v'\|} \leq 1,$$

onde

$$(111) \quad |\langle v | v' \rangle| \leq \|v\| \|v'\|,$$

que é conhecida como *desigualdade de Cauchy-Schwartz*. Desta desigualdade podemos facilmente deduzir uma versão da *desigualdade triangular* para vetores do \mathbb{R}^n . Para isso, observe que, pela distributividade do produto interno,

$$\|v + v'\|^2 = \langle v + v' | v + v' \rangle = \langle v | v \rangle + \langle v' | v' \rangle + 2\langle v | v' \rangle = \|v\|^2 + \|v'\|^2 + 2\langle v | v' \rangle.$$

Aplicando a isto a desigualdade de Cauchy-Schwartz, obtemos

$$\|v + v'\|^2 \leq \|v\|^2 + \|v'\|^2 + 2\|v\| \|v'\|.$$

Como o lado direito desta última desigualdade é igual a $(\|v\| + \|v'\|)^2$, basta extrair a raiz quadrada e teremos

$$\|v + v'\| \leq \|v\| + \|v'\|,$$

que é a versão da desigualdade triangular que esperávamos obter. Esta é a apenas a versão mais recente com que você se deparou de uma desigualdade que vem lhe acompanhando há anos e que é conhecida dos matemáticos há milênios. O nome desigualdade triangular refere-se ao fato de que a versão plana desta desigualdade equivale a dizer que o lado maior de um triângulo é menor ou igual à soma dos lados menores. Esta forma da desigualdade aparece como a proposição 20 do livro I dos Elementos de Euclides.

5.5. Decomposição QR. O algoritmo de Gram-Schmidt tem uma importante interpretação matricial. Suponhamos que, aplicando este algoritmo a um conjunto linearmente independente qualquer

$$\beta = \{v_1, \dots, v_m\}$$

do \mathbb{R}^n , obtemos a base ortonormal

$$\beta^* = \{u_1, \dots, u_m\}.$$

Logo, por construção, u_j é a normalização do vetor w_j definido recursivamente por

$$w_1 = v_1 \text{ e } w_j = v_j - \sum_{i=1}^{j-1} \langle v_j | u_i \rangle u_i.$$

Escrevendo

$$(112) \quad r_{ij} = \begin{cases} \langle v_j | u_i \rangle & \text{se } i \neq j; \\ \|w_i\| & \text{se } i = j, \end{cases}$$

temos que

$$v_1 = w_1 \text{ e } v_j = w_j + \sum_{i=1}^{j-1} r_{ij} u_i;$$

onde

$$(113) \quad v_1 = r_{11} u_1 \text{ e } v_j = r_{jj} u_j + \sum_{i=1}^{j-1} r_{ij} u_i,$$

quando $1 \leq j \leq m$. Portanto, se Q for a matriz cujas colunas são os vetores u_j e R for a matriz cujas entradas são os r_{ij} , teremos

$$QR = \left[\begin{array}{ccc|c} & \cdots & & \\ u_1 & \cdots & u_m & \\ & \cdots & & \end{array} \right] \left[\begin{array}{ccccc} r_{1,1} & r_{1,2} & \cdots & r_{1,n-1} & r_{1,n} \\ 0 & r_{2,2} & \cdots & r_{2,n-1} & r_{2,n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & r_{m,n} \end{array} \right].$$

Efetuando o produto, verificamos que a k -ésima coluna de QR é

$$r_{1,k} u_1 + \cdots + r_{m,k} u_m.$$

em que estamos identificando, como usual, os vetores com as matrizes coluna correspondentes. Mas, por (113), isto corresponde à matriz coluna v_k . Portanto, se denotarmos por A a matriz cujas colunas são os vetores v_1, \dots, v_m , teremos

$$A = QR.$$

Por exemplo, dispondo os vetores da base β do hiperplano no exemplo do artigo 5.3 como colunas de uma matriz A , teremos que

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

Mas os cálculos que fizemos ao aplicar o algoritmo de Gram-Schmidt a β mostram que A pode ser decomposta como produto de

$$(114) \quad Q = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{42} \\ 1/\sqrt{2} & -1/\sqrt{3} & -1/\sqrt{42} \\ 0 & 1/\sqrt{3} & -2/\sqrt{42} \\ 0 & 0 & 6/\sqrt{42} \end{bmatrix} \text{ por } R = \begin{bmatrix} \sqrt{2} & \sqrt{2} & 1/\sqrt{2} \\ 0 & \sqrt{3} & 1/\sqrt{3} \\ 0 & 0 & \sqrt{42}/6 \end{bmatrix},$$

em que as colunas de Q são os vetores de β^* e as entradas de R são definidas por (112).

Um caso especial muito importante é aquele em que A é uma matriz $n \times n$ cujas colunas são linearmente independentes. Neste caso, as colunas de Q formarão uma base ortonormal do \mathbb{R}^n . Logo, se

$$Q = \begin{bmatrix} | & \cdots & | \\ u_1 & \cdots & u_n \\ | & \cdots & | \end{bmatrix}$$

então

$$Q^t Q = \begin{bmatrix} - & u_1^t & - \\ \vdots & \vdots & \vdots \\ - & u_n^t & - \end{bmatrix} \begin{bmatrix} | & \cdots & | \\ u_1 & \cdots & u_n \\ | & \cdots & | \end{bmatrix} = \begin{bmatrix} u_1^t u_1 & u_1^t u_2 & \cdots & u_1^t u_n \\ \vdots & \ddots & & \vdots \\ u_n^t u_1 & u_n^t u_2 & \cdots & u_n^t u_n \end{bmatrix}.$$

Como

$$u_i^t u_j = \langle u_i | u_j \rangle = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j, \end{cases}$$

segue-se que

$$Q^t Q = I.$$

Como vimos no capítulo 1, matrizes com esta propriedade são chamadas de *ortogonais*. Portanto, o cálculo acima mostra que uma matriz $n \times n$ cujas colunas formam uma base ortonormal do \mathbb{R}^n é ortogonal.

A decomposição QR nos dá uma nova maneira de resolver sistemas lineares determinados. Suponhamos que A seja a matriz do sistema, X a matriz das variáveis e b a matriz

dos termos constantes, como definidas no artigo 1.1. Assim, o sistema pode ser escrito na forma

$$AX = b.$$

Se $A = QR$ for a decomposição QR de A , então

$$QRX = b.$$

Mas, pelo que vimos acima, quando o sistema é determinado a matriz Q é ortogonal. Logo, multiplicando ambos os membros da equação anterior, à esquerda, pela transposta de Q , obtemos

$$RX = Q^t(QRX) = Q^t b.$$

Como R é triangular superior, este último sistema pode ser resolvido por substituição.

Vejamos como utilizar este método para resolver o sistema linear

$$(115) \quad \begin{cases} x_1 + x_2 + 2x_3 &= 1, \\ 2x_2 + x_3 &= 2, \\ -1x_1 - x_2 - x_3 &= 5, \end{cases}$$

que tem

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 2 & 1 \\ -1 & -1 & -1 \end{bmatrix}, \quad X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad \text{e} \quad b = \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}$$

como matrizes do sistema, das variáveis e dos termos constantes, respectivamente. Como os vetores

$$v_1 = (1, 0, -1), \quad v_2 = (1, 2, -1) \quad \text{e} \quad v_3 = (2, 1, -1),$$

correspondentes às colunas de A , são linearmente independentes, eles formam uma base do \mathbb{R}^3 . A aplicação do método de Gram-Schmidt a estes vetores, começa com a normalização

$$u_1 = \frac{1}{\sqrt{2}}(1, 0, -1)$$

do vetor v_1 . Em seguida, calculamos

$$r_{12} = \langle u_1 | v_2 \rangle = \frac{2}{\sqrt{2}} = \sqrt{2};$$

de modo que u_2 será a normalização de

$$v_2 - r_{12}u_1 = (0, 2, 0);$$

isto é,

$$u_2 = (0, 1, 0).$$

Finalmente,

$$r_{13} = \langle u_1 | v_3 \rangle = \frac{3}{\sqrt{2}} \quad \text{e} \quad r_{23} = \langle u_2 | v_3 \rangle = 1;$$

onde

$$v_3 - r_{13}u_1 - r_{23}u_2 = \frac{1}{2}(1, 0, 1)$$

cuja normalização é

$$u_3 = \frac{1}{\sqrt{2}}(1, 0, 1).$$

Portanto,

$$Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \quad \text{e} \quad R = \begin{bmatrix} \sqrt{2} & \sqrt{2} & \frac{3}{\sqrt{2}} \\ 0 & 2 & 1 \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Logo, (115) tem as mesmas soluções que $RX = Q^t b$, que, neste exemplo, é igual a

$$\begin{bmatrix} \sqrt{2} & \sqrt{2} & \frac{3}{\sqrt{2}} \\ 0 & 2 & 1 \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix} = \begin{bmatrix} -2\sqrt{2} \\ 2 \\ 3\sqrt{2} \end{bmatrix}.$$

Resolvendo, por substituição, o sistema

$$\left\{ \begin{array}{l} \sqrt{2}x_1 + \sqrt{2}x_2 + \frac{3x_3}{\sqrt{2}} = -2\sqrt{2} \\ 2x_2 + x_3 = 2 \\ \frac{x_3}{\sqrt{2}} = 3\sqrt{2} \end{array} \right.$$

correspondente à equação matricial acima, obtemos

$$x_1 = -9, \quad x_2 = -2 \quad \text{e} \quad x_3 = 6.$$

 À primeira vista pode parecer muito estranho que alguém considere resolver um sistema linear utilizando a decomposição QR. Afinal de contas, mesmo começando com vetores com coordenadas racionais, as normalizações que o método de Gram-Schmidt requer tendem a produzir radicais, o que complica muito os cálculos. Porém, tudo isso pressupõe que os cálculos necessários para resolver os sistemas possam ser feitos de maneira exata. Embora seja este o caso neste livro, o mesmo não se dá em aplicações à física, engenharia ou economia, por exemplo, em que todos os cálculos são feitos com números em ponto flutuante. Nestes casos a decomposição QR tem a vantagem de que a propagação do erro é bem compreendida, o que não ocorre com a eliminação gaussiana. Mais detalhes podem ser encontrados em [7, Lecture 16, pp. 114–120].

6. Operações com subespaços

Subespaços são conjuntos de vetores, de modo que podemos aplicar a eles qualquer uma das operações de conjuntos que conhecemos. Contudo, do ponto de vista da álgebra linear, só são interessantes aquelas operações que, quando aplicadas a subespaços, produzem novos subespaços. Nesta seção investigamos quais das operações de conjunto satisfazem este requisito e veremos como adaptar aquelas que não o satisfazem.

6.1. Interseção, união e complementar. De todas as operações de conjunto, a mais fácil de analisar é a interseção. Suponhamos que U e U' sejam subespaços do \mathbb{R}^n . Como $0 \in U$ e $0 \in U'$, temos que $0 \in U \cap U'$, o que garante que a interseção de dois subespaços nunca é o conjunto vazio. Por outro lado, se v e w são vetores de $U \cap U'$, então pertencem tanto a U , quanto a U' , donde

$$v + w \in U \text{ e } v + w \in U',$$

já que ambos são subespaços. Logo, $v + w \in U \cap U'$. Finalmente, se $v \in U \cap U'$ e $\lambda \in \mathbb{R}$, então

$$\lambda v \in U \text{ e } \lambda v \in U',$$

onde $\lambda v \in U \cap U'$. Portanto a interseção de dois subespaços, de um mesmo espaço \mathbb{R}^n , também é um subespaço do \mathbb{R}^n .

É muito fácil calcular a interseção de dois subespaços do \mathbb{R}^n que estão descritos como conjuntos solução de sistemas homogêneos. Para ver como fazer isto, digamos que

$$S_A = \{v \in \mathbb{R}^n \mid Av = 0\} \text{ e } S_B = \{v \in \mathbb{R}^n \mid Bv = 0\},$$

em que A e B são matrizes de tamanhos $k \times n$ e $\ell \times n$, respectivamente. Um vetor v pertence a S_A e S_B se satisfaz simultaneamente à condição para estar em S_A e à condição para estar em S_B ; isto é,

$$S_A \cap S_B = \{v \in \mathbb{R}^n \mid Av = 0 \text{ e } Bv = 0\}.$$

Em outras palavras, v será solução de um sistema com $k + \ell$ equações, obtidas pela reunião das equações do sistema $AX = 0$ com o sistema $BX = 0$, em que X é o vetor coluna das variáveis. Por exemplo, a interseção dos subespaços

$$(116) \quad S_A = \{(x, y, z, w) \in \mathbb{R}^4 \mid x + y + z + w = 0\}$$

e

$$(117) \quad S_B = \{(x, y, z, w) \in \mathbb{R}^4 \mid x + 2y = z + 3w = 0\}$$

do \mathbb{R}^4 é igual a

$$S_A \cap S_B = \{(x, y, z, w) \in \mathbb{R}^4 \mid x + y + z + w = x + 2y = z + 3w = 0\}.$$

Resolvendo o sistema, descobrimos que se trata da reta gerada pelo vetor $(4, -2, -3, 1)$.

Menos óbvio é o que acontece com a união: é possível que um subespaço seja igual à uma união de subespaços, mas isto só acontece quando um subespaço está contido em outro. Para um exemplo simples de união que não é subespaço, tome duas retas passando pela origem. Cada uma das retas é um subespaço, mas se somarmos vetores não nulos, um em cada reta, teremos um vetor que não está em uma, nem na outra, como ilustrado na figura 3 para o caso de retas do plano. Portanto, a união de duas retas distintas nunca será um subespaço. Note que, neste exemplo, obteremos um subespaço se considerarmos o conjunto que contém, não apenas a união das retas, mas também todas as somas de um

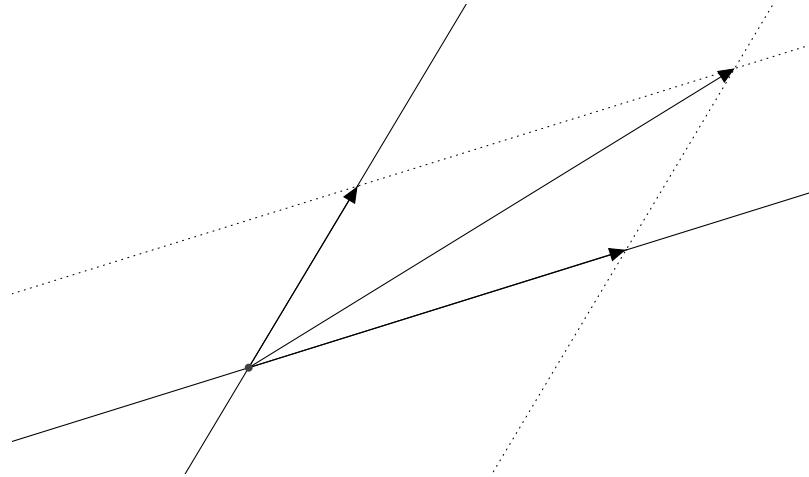


FIGURA 3. A união de duas retas pela origem não é um subespaço do plano

vetor em uma das retas com um vetor na outra reta, porque isto equivale ao subespaço gerado pelos dois vetores. Teremos mais a dizer sobre esta observação no próximo artigo.

A situação do complementar é ainda mais radical, porque não há nenhuma condição sob a qual retorne um subespaço. A razão é que um subespaço tem que conter o vetor nulo, de modo que seu complementar nunca conterá este vetor. Logo, o complementar de um subespaço nunca é um subespaço. Apesar disto, veremos, no artigo 6.3, como definir um subespaço que, no contexto da álgebra linear, desempenha um papel semelhante ao do complementar usual no âmbito da teoria de conjuntos.

6.2. Soma de subespaços. Sejam U e U' dois subespaços do \mathbb{R}^n . As considerações que fizemos sobre a união no artigo anterior, sugerem a seguinte definição. A *soma* de U e U' é o conjunto

$$U + U' = \{u + u' \mid u \in U \text{ e } u' \in U'\};$$

cujos elementos são obtidos somando cada vetor de U com cada vetor de U' . Note que tanto U , quanto U' , estão contidos em $U + U'$ pois o vetor nulo pertence a U e a U' , de modo que $u + 0 \in U + U'$ e $0 + u' \in U + U'$, quaisquer que sejam os vetores $u \in U$ e $u' \in U'$. É fácil mostrar, diretamente da definição, que $U + U'$ é subespaço do \mathbb{R}^n ; não faremos isto aqui porque nossa discussão sobre como achar geradores para $U + U'$ trará, em si, uma demonstração de que este conjunto é um subespaço.

No artigo anterior, vimos que é fácil calcular a interseção de subespaços definidos como soluções de sistemas lineares. A soma de subespaços, por sua vez, é fácil de calcular quando conhecemos geradores para os dois subespaços. Digamos que

$$U = \langle w_1, \dots, w_k \rangle \quad \text{e} \quad U' = \langle w'_1, \dots, w'_{\ell} \rangle.$$

Vimos que U e U' estão ambos contidos em $U + U'$, de modo que

$$w_1, \dots, w_k, w'_1, \dots, w'_\ell \in U + U'.$$

Por outro lado, como

$$a_1 w_1 + \dots + a_k w_k \in U \quad \text{e} \quad a'_1 w'_1 + \dots + a'_\ell w'_\ell \in U',$$

quaisquer que sejam $a_1, \dots, a_k, a'_1, \dots, a'_\ell \in \mathbb{R}$, obtemos, pela definição de $U + U'$, que

$$(a_1 w_1 + \dots + a_k w_k) + (a'_1 w'_1 + \dots + a'_\ell w'_\ell) \in U + U';$$

que equivale a dizer que

$$(118) \quad \langle w_1, \dots, w_k, w'_1, \dots, w'_\ell \rangle \subset U + U'.$$

Reciprocamente, se $u \in U$ e $u' \in U'$, podemos escrevê-los como combinações lineares de w_1, \dots, w_k e de w'_1, \dots, w'_ℓ , respectivamente; o que mostra que

$$u + u' \in \langle w_1, \dots, w_k, w'_1, \dots, w'_\ell \rangle$$

e prova a inclusão oposta à obtida em (118). Portanto, $U + U'$ é igual ao subespaço gerado pela união de um conjunto de geradores de U com um conjunto de geradores de U' . Em particular, provamos que $U + U'$ é, de fato um subespaço do \mathbb{R}^n .

Vamos aplicar o que aprendemos aos subespaços S_A e S_B definidos em (116) e (117). Antes, porém, precisamos encontrar geradores para estes subespaços. Como ambos os sistemas são definidos por matrizes em forma escada, isto é bastante fácil de fazer. No caso de S_A , temos que $(x, y, z, w) \in S_A$ se, e somente se,

$$(x, y, z, w) = (-y - z - w, y, z, w) = y(-1, 1, 0, 0) + z(-1, 0, 1, 0) - w(-1, 0, 0, 1);$$

do que podemos concluir que

$$S_A = \langle (-1, 1, 0, 0), (-1, 0, 1, 0), (-1, 0, 0, 1) \rangle.$$

Por outro lado, $(x, y, z, w) \in S_B$ se, e somente se,

$$(x, y, z, w) = (-2y, y, -3w, w) = y(-2, 1, 0, 0) + w(0, 0, -3, 1);$$

onde

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

Portanto,

$$S_A + S_B = \langle (-1, 1, 0, 0), (-1, 0, 1, 0), (-1, 0, 0, 1), (-2, 1, 0, 0), (0, 0, -3, 1) \rangle.$$

Como a forma escada desta matriz é

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

verificamos que os vetores

$$(-1, 1, 0, 0), (0, -1, 1, 0), (0, 0, -1, 1), (0, 0, 0, -1)$$

formam uma base de $S_A + S_B$. Logo, pelo corolário 4.4, $S_A + S_B = \mathbb{R}^4$.

Há um caso especial da soma de subespaços que é tão importante que tem seu próprio nome e símbolo. Sejam U , U' e W subespaços do \mathbb{R}^n . Dizemos que W é *soma direta* de U e U' se

$$W = U + U' \quad \text{e} \quad U \cap U' = \{0\}.$$

Quando isto acontece, escrevemos $W = U \oplus U'$. Por exemplo, se ℓ e ℓ' são duas retas distintas do plano, então

$$\mathbb{R}^2 = \ell \oplus \ell'.$$

A proposição abaixo é uma das razões pela qual a soma direta é tão importante.

PROPOSIÇÃO 6.1. *Sejam U , U' e W subespaços do \mathbb{R}^n . Se $W = U \oplus U'$, então a união de uma base de U com uma base de U' produz uma base de W ; de modo que $\dim(W) = \dim(U) + \dim(U')$.*

DEMONSTRAÇÃO. Sejam

$$\beta = \{u_1, \dots, u_k\} \quad \text{e} \quad \beta' = \{u'_1, \dots, u'_m\}$$

bases de U e U' , respectivamente. Como W é a soma entre U e U' , é claro que $\beta \cup \beta'$ gera W . Resta mostrar que $\beta \cup \beta'$ é linearmente independente. Para isso, considere a combinação linear

$$a_1u_1 + \cdots + a_ku_k + a'_1u'_1 + \cdots + a'_mu'_m = 0,$$

que podemos reescrever na forma

$$a_1u_1 + \cdots + a_ku_k = -a'_1u'_1 - \cdots - a'_mu'_m.$$

O lado direito desta equação é uma combinação linear de vetores da base β de U e, portanto, pertence a U . Analogamente, o lado esquerdo é uma combinação linear de vetores de β' e, portanto, pertence a U' . Logo,

$$a_1u_1 + \cdots + a_ku_k = -a'_1u'_1 - \cdots - a'_mu'_m \in U \cap U' = \{0\};$$

onde

$$a_1u_1 + \cdots + a_ku_k = -a'_1u'_1 - \cdots - a'_mu'_m = 0.$$

Finalmente, como os vetores de β e de β' são linearmente independentes,

$$a_1 = \cdots = a_k = a'_1 = \cdots = a'_m = 0$$

o que mostra que $\beta \cup \beta'$ é um conjunto de vetores linearmente independentes. Como β e β' não têm vetores em comum, temos que

$$\dim(W) = \#(\beta \cup \beta') = \#(\beta) + \#(\beta') = \dim(U) + \dim(U'),$$

o que prova a afirmação sobre as dimensões. □

 É importante notar que o resultado desta proposição é, em geral, falso quando a condição $U \cap U' = \{0\}$ não é satisfeita. Por um exemplo, quando U e U' são dois planos distintos do \mathbb{R}^3 . A não ser que as bases que escolhamos para estes planos tenham um vetor em comum, sua união não pode ser uma base de $U + U'$ porque, pelo lema da troca, nenhum subconjunto linearmente independente do \mathbb{R}^3 pode ter mais de três elementos.

6.3. Subespaço complementar. No artigo anterior, vimos que o papel da união é desempenhado, em álgebra linear, pela soma. Neste artigo queremos descobrir qual espaço deveria corresponder ao complementar de um conjunto e como podemos calculá-lo.

Dados dois conjuntos $Y \subset X$, o **complementar** Y^c de Y em X é o conjunto formado pelos elementos de X que não estão em Y . Portanto,

$$(119) \quad Y \cup Y^c = X \quad \text{e} \quad Y \cap Y^c = \emptyset.$$

Usaremos estas duas equações como ponto de partida para determinar como deveria ser definido o complementar \bar{U} de um subespaço U do \mathbb{R}^n . Como já sabemos que o papel da união de conjuntos é desempenhado pela soma de subespaços, concluímos que a primeira equação em (119) deve ser substituída por $U + \bar{U} = \mathbb{R}^n$. Por outro lado, como todo subespaço do \mathbb{R}^n contém o vetor nulo, a menor interseção possível entre U e \bar{U} é $\{0\}$. Isto sugere que deveríamos tomar $U \cap \bar{U} = \{0\}$ como substituta da segunda equação de (119).

Resta-nos oficializar a definição sugerida pelas considerações feitas no parágrafo anterior. Sejam U e \bar{U} dois subespaços do \mathbb{R}^n . Se

$$(120) \quad U + \bar{U} = \mathbb{R}^n \quad \text{e} \quad U \cap \bar{U} = \{0\}$$

diremos que U e \bar{U} são subespaços *complementares*. Temos, assim, que o \mathbb{R}^n é a soma direta de U e \bar{U} . Frequentemente nos referiremos a \bar{U} como sendo um *subespaço complementar* de U . Note o uso do artigo indefinido: como veremos em breve, qualquer subespaço do \mathbb{R}^n , diferente de $\{0\}$ e do \mathbb{R}^n inteiro, tem infinitos complementares.

Agora que temos uma definição para subespaços, precisamos descobrir como proceder para achar um complementar para um dado subespaço U do \mathbb{R}^n . Digamos que conhecemos uma base $\beta = \{u_1, \dots, u_k\}$ de U . Sabemos, pelo teorema 4.3, que é possível acrescentar vetores w_{k+1}, \dots, w_n a β de modo a obter uma base do \mathbb{R}^n . É, então, razoável esperar que o subespaço \bar{U} gerado por esses vetores seja um complementar de U . Para confirmar isto precisamos mostrar que o subespaço \bar{U} , assim definido, satisfaz as duas equações em (120).

Em primeiro lugar, $U + \bar{U}$ contém tanto U , quanto \bar{U} , de modo que

$$u_1, \dots, u_k, w_{k+1}, \dots, w_n \in U + \bar{U}.$$

Como estes vetores formam uma base do \mathbb{R}^n , é possível gerar todos os vetores deste último espaço a partir deles; logo $U + \bar{U} = \mathbb{R}^n$, como esperávamos mostrar.

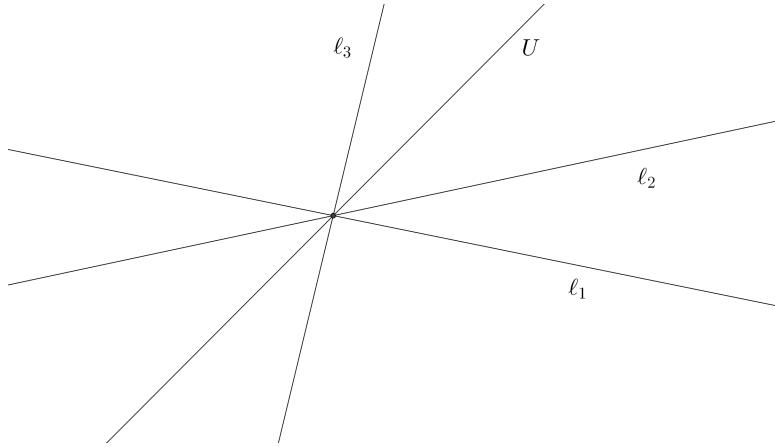


FIGURA 4. Alguns complementares do subespaço U .

A segunda equação em (120) exige um pouco mais de esforço. Digamos que

$$v \in U \cap \overline{U}.$$

Como u_1, \dots, u_k é uma base de U e $v \in U$, existem números reais a_1, \dots, a_k tais que

$$v = a_1 u_1 + \cdots + a_k u_k.$$

Mas, v também está em \overline{U} , de modo que

$$v = a_{k+1} w_{k+1} + \cdots + a_n w_n.$$

para alguma escolha de $a_{k+1}, \dots, a_n \in \mathbb{R}$. Subtraindo a segunda expressão para v da primeira, obtemos

$$a_1 u_1 + \cdots + a_k u_k - a_{k+1} w_{k+1} - \cdots - a_n w_n = 0.$$

Porém, por construção, os vetores

$$u_1, \dots, u_k, w_{k+1}, \dots, w_n$$

formam uma base do \mathbb{R}^n e, portanto, são linearmente independentes. Logo,

$$a_1 = \cdots = a_k = a_{k+1} = \cdots = a_n = 0.$$

Consequentemente, $v = 0$, o que nos permite afirmar que o vetor nulo é o único elemento de $U \cap \overline{U}$, provando assim que a segunda equação em (120) também é verificada.

Usando a construção acima é fácil dar exemplos de uma infinidade de complementares diferentes para um dado espaço. Para que seja possível, literalmente, ver os complementares, vamos considerar o caso em que U é uma reta do \mathbb{R}^2 . Neste caso qualquer reta diferente de U é um complementar de U . As retas ℓ_1 , ℓ_2 e ℓ_3 da figura 4 ilustram alguns subespaços complementares da reta $x - y = 0$.

Para um exemplo mais elaborado, construiremos o complementar do subespaço U do \mathbb{R}^5 gerado pelos vetores do conjunto

$$G = \{(1, 0, -1, 2, 3), (-3, 0, 3, -5, -10), (-1, 0, 1, -4, -2)\}.$$

Escalonando a matriz $A(G)$, cujas linhas são os geradores de U , obtemos

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

Portanto, pela proposição 4.2,

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

é uma base de U . Contudo, como se tratam de vetores do \mathbb{R}^5 , podemos acrescentar duas linhas à matriz sem que ela deixe de ser uma matriz escada. Naturalmente os vetores a serem acrescentados precisam estar entre a primeira e a segunda linhas e entre a segunda e a terceira linhas, para que a forma escada seja mantida. Quaisquer dois vetores podem ser acrescentados, desde que a forma escada seja preservada. Assim, podemos completar β com quaisquer vetores da forma

$$v = (0, a_1, a_2, a_3, a_4) \quad \text{e} \quad w = (0, 0, b_1, b_2, b_3),$$

em que $a_1, a_2, a_3, a_4, b_1, b_2 \in \mathbb{R}$, desde que a_1 e b_1 sejam ambos diferentes de zero. Como há uma infinidade de escolhas possíveis para os a 's e b 's, há uma infinidade de complementares possíveis para o subespaço U . Por exemplo, os subespaços

$$\bar{U}_n = \langle (0, 1, 0, n, 0), (0, 0, 1, 0, 0) \rangle$$

são todos complementares diferentes de U . No próximo artigo veremos que há uma outra maneira, também muito simples, de achar um complementar que é bastante conveniente quando o subespaço é definido como conjunto solução de um sistema homogêneo.

6.4. Complemento ortogonal. Uma outra maneira de construir subespaços complementares utiliza o produto interno. Se U é um subespaço do \mathbb{R}^n , definimos seu *complemento ortogonal* como o conjunto dos vetores do \mathbb{R}^n que são ortogonais a todos os vetores de U ; isto é,

$$U^\perp = \{v \in \mathbb{R}^n \mid \langle v | u \rangle = 0 \text{ para todo } u \in U\}.$$

Um problema com esta definição é que, aparentemente, requer que testemos se um vetor é ortogonal aos infinitos vetores de U . É em situações como esta que fica claro porque é tão importante saber que todo subespaço do \mathbb{R}^n admite um conjunto finito de geradores. De fato, se

$$u_1, \dots, u_m$$

são geradores de U e $v \in \mathbb{R}^n$ é ortogonal a cada um desses vetores, então

$$\langle v | a_1 u_1 + \dots + a_m u_m \rangle = a_1 \langle v | u_1 \rangle + \dots + a_m \langle v | u_m \rangle = 0.$$

Portanto, se um vetor é ortogonal a todos os vetores de um conjunto de geradores de U , então é ortogonal a todos os vetores de U . Como a recíproca desta afirmação é claramente verdadeira, podemos concluir que, para achar o complemento ortogonal de U , basta achar todos os vetores ortogonais a um conjunto de geradores de U .

Por exemplo, o complemento ortogonal de

$$(121) \quad W = \langle (1, 0, 2, 0), (0, 2, 3, -1) \rangle$$

é o conjunto dos vetores (x_1, x_2, x_3, x_4) cujo produto interno com $(1, 0, 2, 0)$ e $(0, 1, 3, -1)$ é nulo. Mas,

$$\begin{aligned} \langle (1, 0, 2, 0) | (x_1, x_2, x_3, x_4) \rangle &= x_1 + 2x_3 \text{ e} \\ \langle (0, 2, 3, -1) | (x_1, x_2, x_3, x_4) \rangle &= 2x_2 + 3x_3 - x_4, \end{aligned}$$

de modo que W^\perp é o conjunto solução do sistema

$$x_1 + 2x_3 = 2x_2 + 3x_3 - x_4 = 0.$$

Logo, $(x_1, x_2, x_3, x_4) \in W^\perp$ se, e somente se,

$$(x_1, x_2, x_3, x_4) = (-2x_3, x_2, x_3, 2x_2 + 3x_3) = x_2(0, 1, 0, 2) + x_3(-2, 0, 1, 3).$$

Portanto, pela proposição 4.2,

$$\{(1, 0, 2, 0), (0, 2, 3, -1)\} \quad \text{e} \quad \{(0, 1, 0, 2), (-2, 0, 1, 3)\}$$

são bases dos subespaços W e W^\perp , respectivamente. Seja G o conjunto formado por estes quatro vetores. Aplicando eliminação à matriz $A(G)$ descobrimos que sua forma escada não contém nenhuma linha nula. Logo, utilizando a proposição 4.2 e o corolário 4.4, podemos concluir que G é uma base do \mathbb{R}^4 . Portanto, W^\perp é um subespaço complementar de W . Como mostra a próxima proposição, o mesmo se aplica ao complemento ortogonal de qualquer subespaço do \mathbb{R}^n .

PROPOSIÇÃO 6.2. *Se U é um subespaço do \mathbb{R}^n , então:*

- (a) U^\perp é um complementar de U ;
- (b) $\mathbb{R}^n = U \oplus U^\perp$;
- (c) $(U^\perp)^\perp = U$.

DEMONSTRAÇÃO. Para provar que U^\perp é um complementar de U precisamos mostrar que

$$(122) \quad U \cap U^\perp = \{0\} \quad \text{e que} \quad U + U^\perp = \mathbb{R}^n.$$

Mas, se v é um vetor na interseção $U \cap U^\perp$ então, pela definição do complemento ortogonal

$$\langle v | v \rangle = 0.$$

Como o produto interno é positivo definido, temos disto que $v = 0$, o que prova a primeira equação em (122). Para provar a segunda equação, suponha que

$$\beta = \{u_1, \dots, u_k\}$$

é uma base de U cuja matriz $A(\beta)$ está em forma escada. Então, U^\perp é o conjunto solução do sistema homogêneo $A(\beta) \cdot X = 0$. Portanto, pela proposição 4.5,

$$(123) \quad \dim(U^\perp) = n - \dim(U).$$

Mas, pela proposição 6.1 a união de uma base de U com uma base de U^\perp é uma base de $U \oplus U^\perp$. Logo, por (123),

$$\dim(U \oplus U^\perp) = n.$$

Porém, pelo corolário 4.4, isto só pode acontecer se

$$U^\perp \oplus U = \mathbb{R}^n,$$

o que prova a segunda equação em (122). O item (b) é consequência imediata do item (a). Para provar (c), note que, pela proposição 6.1 e o item (b), temos

$$n = \dim(U^\perp \oplus (U^\perp)^\perp) = \dim(U^\perp) + \dim((U^\perp)^\perp);$$

onde

$$\dim((U^\perp)^\perp) = n - \dim(U^\perp) = \dim(U).$$

Mas,

$$U \subset (U^\perp)^\perp,$$

pois os vetores de U^\perp são, por construção, ortogonais a todos os vetores de U . Como estes dois subespaços têm a mesma dimensão, segue-se do corolário 4.4 que têm que ser iguais. \square

O que aprendemos sobre o complemento ortogonal nos permite construir um sistema linear cujo conjunto solução é um dado subespaço U do \mathbb{R}^n . Se

$$\beta = \{u_1, \dots, u_k\}$$

for uma base de U^\perp , então,

$$\langle u_1 | v \rangle = \dots = \langle u_k | v \rangle = 0,$$

para todo $v \in U$. Mas isto equivale a dizer que U é o conjunto solução do sistema linear $A(\beta)X = 0$, em que X é o vetor coluna cujas entradas são as variáveis x_1, \dots, x_n . A aplicação mais simples deste fato é à caracterização dos hiperplano em termos de sua dimensão.

PROPOSIÇÃO 6.3. *Um subespaço do \mathbb{R}^n é um hiperplano se tem dimensão igual a $n - 1$.*

DEMONSTRAÇÃO. Já vimos no corolário 4.6 que todo hiperplano tem dimensão $n - 1$. Falta-nos, apenas, provar que todo subespaço S de dimensão $n - 1$ do \mathbb{R}^n tem que ser um hiperplano. Mas, pelas proposições 6.1 e 6.2,

$$\dim(S^\perp) = n - \dim(S) = n - (n - 1) = 1.$$

Logo, existe um vetor não nulo $v_0 \in \mathbb{R}^n$ tal que $S^\perp = \langle v_0 \rangle$. Portanto, pelo argumento acima,

$$S = \{w \in \mathbb{R}^n \mid \langle w \mid v_0 \rangle = 0\}.$$

Se

$$v_0 = (a_1, \dots, a_n),$$

então

$$a_1x_1 + \dots + a_nx_n = \langle (x_1, \dots, x_n) \mid (a_1, \dots, a_n) \rangle = 0$$

nos dá a equação de S , provando que se trata de um hiperplano. \square

Vejamos uma outra aplicação da mesma estratégia, desta vez para achar o sistema homogêneo do qual o subespaço W , definido em (121), é o conjunto solução. Já vimos que

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

é uma base do complemento ortogonal de W . Como $W = (W^\perp)^\perp$, pela proposição 6.2, segue-se que W é o conjunto solução do sistema

$$\begin{bmatrix} 0 & 1 & 0 & 2 \\ -2 & 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

que equivale a

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

Encerraremos o capítulo com um resultado que mostra a equivalência das duas maneiras que usamos para definir um subespaço. Em primeiro lugar, a proposição 4.5 nos permite afirmar que todo conjunto solução de sistema homogêneo admite uma base e, portanto, um conjunto finito de geradores. Por outro lado, pelo teorema 4.3, um subespaço U do \mathbb{R}^n sempre admite uma base. Podemos, então, usar esta base para achar o complemento ortogonal U^\perp de U . Finalmente, o procedimento acima nos dá uma maneira de determinar, a partir de uma base de U^\perp , um sistema homogêneo, do qual U é solução. Com isso demonstramos o resultado abaixo.

TEOREMA 6.4. *Todo subespaço do \mathbb{R}^n pode ser definido por um conjunto de geradores ou como o conjunto solução de um sistema homogêneo.*

Exercícios

1. Quais dos seguintes subconjuntos do \mathbb{R}^n são subespaços? Justifique a sua resposta detalhadamente.
 - (a) $\{(x, y) \in \mathbb{R}^2 \mid 2x - y = 0\}$;
 - (b) $\{(x, y, z) \in \mathbb{R}^3 \mid x - y + 9z + 4 = 0\}$;
 - (c) $\{(x, y) \in \mathbb{R}^2 \mid 2x - y^2 = 0\}$;
 - (d) $\{(x, y, z, w) \in \mathbb{R}^4 \mid 2x - y = 4x - 3y + 7z + w = 0\}$.
2. Quais são os sistemas do exercício 3 da página 109 cujos conjuntos solução são subespaços? Justifique detalhadamente a sua resposta.
3. Escreva o vetor v como combinação linear dos vetores do conjunto G , para cada um dos exemplos abaixo.
 - (a) $v = (2, 8)$ e $G = \{(1, -1), (3, 2)\}$;
 - (b) $v = (0, 1)$ e $G = \{(3, 2), (2, 2)\}$;
 - (c) $v = (2, 1, 3)$ e $G = \{(1, 1, 1), (1, 1, -1), (1, -4, 5)\}$;
 - (d) $v = (1, -1, 4)$ e $G = \{(1, -1, 2), (1, -3, 4), (-1, 3, 2)\}$;
 - (e) $v = (2, 1, 3, 2)$ e $G = \{(1, 0, 0, 1), (0, 1, 2, 0), (0, 1, -1, 0)\}$;
 - (f) $v = (1, 1, 1, 1)$ e $G = \{(2, 3, 1, 1), (5, 6, 1, 1), (1, 2, 1, 1)\}$.
4. Determine um conjunto finito de geradores para cada um dos seguintes subespaços:
 - (a) $\{(x, y, z) \in \mathbb{R}^3 \mid x + y - 2z = 0\}$ em \mathbb{R}^3 ;
 - (b) $\{(x, y, z) \in \mathbb{R}^3 \mid x - y + z = y - z = 0\}$ em \mathbb{R}^3 ;
 - (c) $\{(x, y, z, w) \in \mathbb{R}^4 \mid x + z = y - w = 0\}$ em \mathbb{R}^4 ;
 - (d) $\{(x, y, z, w) \in \mathbb{R}^4 \mid x + y - z = x - y + z - w = x - z + w = 0\}$ em \mathbb{R}^4 ;
 - (e) $\{(x, y, z, w) \in \mathbb{R}^4 \mid x + y + z + w = 0\}$ em \mathbb{R}^4 .
5. Determine um conjunto finito de geradores para o conjunto solução do sistema homogêneo associado a cada um dos sistemas do exercício 3 da página 109.
6. Sejam U e U' subespaços de um espaços do \mathbb{R}^n .
 - (a) Mostre, diretamente da definição, que $U + U'$ é um subespaço do \mathbb{R}^n .
 - (b) Mostre que todo subespaço do \mathbb{R}^n que contém U e U' também contém $U + U'$.
7. Quais dos subconjuntos abaixo são linearmente dependentes e quais são linearmente independentes? Justifique a sua resposta.
 - (a) $\{(1, 0, 1), (1, 3, 2), (1, 1, -1)\}$ em \mathbb{R}^3 ;
 - (b) $\{(1, 0, 1, -1), (1, 1, 0, 1), (0, 1, -1, 2), (0, 2, 2, -4)\}$ em \mathbb{R}^4 ;
 - (c) $\{(1, 1, -1), (2, 3, -1), (3, 1, -5)\}$ em \mathbb{R}^3 ;
 - (d) $\{(3, 1, 2, 3), (2, 1, 3, 2), (1, 3, 2, 1)\}$ em \mathbb{R}^4 ;
 - (e) $\{(1, 2, 3), (5, 6, 7), (8, 1, 2), (1, 0, 1)\}$ em \mathbb{R}^3 ;
 - (f) $\{(1, 0, 0, 0), (1, 1, 0, 0), (1, 1, 1, 0), (1, 1, 1, 1)\}$ em \mathbb{R}^4 ;
 - (g) $\{(1, 0, 0, 0), (1, 1, 1, 1), (1, 3, 4, 1), (6, 4, 5, 2)\}$ em \mathbb{R}^4 .

8. Quais dos subconjuntos abaixo são linearmente dependentes e quais são linearmente independentes? No caso daqueles que são linearmente dependentes, escreva cada um dos vetores do conjunto (quando possível), como combinação linear dos demais.
- $\{(1, 2, -3, 1), (2, 0, 4, 1), (5, -4, 14, -3)\}$ em \mathbb{R}^4 ;
 - $\{(2, 1, 2, 1), (6, 3, 6, 3), (5, 1, 4, 3)\}$ em \mathbb{R}^4 ;
 - $\{(1, 1, 0), (1, 0, 1), (0, 1, 1)\}$ em \mathbb{R}^3 ;
 - $\{(1, 0, 0, 1), (1, 1, 0, 1), (1, 1, 1, 1), (1, 0, 1, 1), (1, 0, 1, 0)\}$ em \mathbb{R}^4 .
9. Sejam $u, v, w \in \mathbb{R}^n$. Sabendo-se que estes vetores são linearmente independentes, determine se o mesmo ocorre com os seguintes vetores:
- $u + v, v + w$ e $u + w$;
 - $u - v, v - w$ e $w - u$.
10. Calcule a base e a dimensão de cada um dos seguintes subespaços:
- $\langle(1, 2, 3), (1, 1, 1), (-1, 0, 1), (1, 4, 7)\rangle$ em \mathbb{R}^3 ;
 - $\langle(1, -1, 0, 2), (3, 5, 2, -4), (5, 3, 2, 0)\rangle$ em \mathbb{R}^4 ;
 - $\langle(1, 1, 2, 3, 6), (1, 2, 3, 1, 6), (1, 0, 1, 0, 9)\rangle$ em \mathbb{R}^5 ;
 - $\{(x, y, z) \in \mathbb{R}^3 \mid x + y - 2z = 0\}$ em \mathbb{R}^3 ;
 - $\{(x, y, z) \in \mathbb{R}^3 \mid x - y + z = y - z = 0\}$ em \mathbb{R}^3 ;
 - $\{(x, y, z, w) \in \mathbb{R}^4 \mid x + z = y - w = 0\}$ em \mathbb{R}^4 ;
 - $\{(x, y, z, w) \in \mathbb{R}^4 \mid x + y - z = x - y + z - w = x - z + w = 0\}$ em \mathbb{R}^4 ;
 - $\{(x, y, z, w) \in \mathbb{R}^4 \mid x + y + z + w = 0\}$ em \mathbb{R}^4 .
11. Seja $\{v_1, \dots, v_n\} \subset \mathbb{R}^n$ um conjunto de vetores linearmente independentes e considere
- $$S = \{v_1 - v_2, v_1 - v_3, \dots, v_1 - v_n\}.$$
- Quais das seguintes afirmações são verdadeiras:
- o conjunto S é linearmente independente;
 - S é um conjunto de geradores de \mathbb{R}^n ;
 - S é uma base de \mathbb{R}^n .
12. Dados números reais r_1, \dots, r_n defina os polinômios
- $$\ell_i = \frac{(x - r_1) \cdots (x - r_{i-1})(x - r_{i+1}) \cdots (x - r_n)}{(r_i - r_1) \cdots (r_i - r_{i-1})(r_i - r_{i+1}) \cdots (r_i - r_n)}$$
- para cada $1 \leq i \leq n$. Mostre que os polinômios ℓ_1, \dots, ℓ_n definem uma base do espaço dos polinômios na variável x , com coeficientes reais e grau menor ou igual a n .
13. Mostre que os subespaços de \mathbb{R}^4 definidos abaixo são iguais:
- $$U = \langle(1, -1, -3), (3, -2, -8), (2, 1, -3)\rangle;$$
- $$U' = \langle(1, 1, -1), (2, 3, -1), (3, 1, -5)\rangle.$$

14. Calcule uma base e a dimensão do conjunto solução sistema linear

$$\begin{cases} x + 2y - z + w - t = 0 \\ x - 2z - w + t = 0 \\ y + z + w - t = 0 \\ 2x - 6y - 7z - 8w + 8t = 0 \\ 3x + 7y - 2z + 4w - 4t = 0. \end{cases}$$

15. Dê exemplo de um sistema linear homogêneo cujo conjunto solução coincide com cada um dos subespaços vetoriais abaixo:

- (a) $\langle(1, 1, -1), (2, 3, -1), (3, 1, -5)\rangle$ em \mathbb{R}^3 ;
- (b) $\langle(1, -1, 0, 0), (1, 0, 1, 1), (0, 1, -1, 0)\rangle$ em \mathbb{R}^4 ;
- (c) $\langle(1, -1, 0, 1, 0), (-1, 1, 0, 0, 1)\rangle$ em \mathbb{R}^5 ;
- (d) $\langle(1, 1, 1, 0), (1, 0, 2, 1), (5, 2, 8, 3)\rangle$ em \mathbb{R}^4 ;
- (e) $\langle(1, 2, 0, 1, 1), (1, 4, 1, 2, 0), (1, 0, 1, 1, 1)\rangle$ em \mathbb{R}^5 .

16. Complete cada um dos seguintes conjuntos para uma base de \mathbb{R}^4 :

- (a) $\{(2, 1, 4, 3), (2, 1, 2, 0)\}$;
- (b) $\{(0, 1, 2, 3), (1, 2, 3, 4), (0, 0, 0, 1)\}$;
- (c) $\{(0, 2, 1, 0), (0, 1, 2, 0)\}$;
- (d) $\{(1, 2, 3, 4), (4, 3, 2, 1)\}$;

17. Considere os seguintes subespaços de \mathbb{R}^4 :

$$U = \{(x, y, z, w) \in \mathbb{R}^4 \mid x + z = y - w = 0\};$$

$$W = \{(x, y, z, w) \in \mathbb{R}^4 \mid x - w = 0\}.$$

Determine:

- (a) bases para U e W ;
- (b) as dimensões de U e W ;
- (c) uma base de \mathbb{R}^4 que contenha uma base de U ;
- (d) uma base de $U + W$ e sua dimensão;
- (e) uma base de $U \cap W$ e sua dimensão;
- (f) um subespaço U' tal que $U + U' = \mathbb{R}^4$.

18. Considere os seguintes subespaços de \mathbb{R}^4 :

$$U = \langle(1, 0, 1, 1), (1, -1, 1, 1), (0, 1, 0, 0)\rangle;$$

$$W = \{(x, y, z, w) \in \mathbb{R}^4 \mid x - y + z - w = 0\}.$$

Determine:

- (a) bases para U e W ;
- (b) as dimensões de U e W ;
- (c) um sistema homogêneo cujo conjunto solução é U ;

- (d) uma base de $U + W$ e sua dimensão;
 (e) uma base de $U \cap W$ e sua dimensão;

19. Considere os seguintes subespaços de \mathbb{R}^4 :

$$U = \{(x, y, z, w) \in \mathbb{R}^4 \mid x - y + z = x + y - z + w = 0\};$$

$$W = \{(x, y, z, w) \in \mathbb{R}^4 \mid y + z - w = x + ky - w = 0\}.$$

- (a) Determine os valores reais de k para os quais $\dim(U \cap W) = 1$.
 (b) Determine os valores reais de k para os quais $\dim(U + W) = 3$.

20. Ache uma base para $U \cap W$ sabendo-se que U e W são os seguintes subespaços de \mathbb{R}^4 :

$$U = \langle(4, -3, 2, 0), (7, 0, 5, 3)\rangle;$$

$$W = \langle(2, -5, 3, 1), (5, -2, 6, 4), (7, -7, 9, 5)\rangle;$$

21. Mostre que se U e W são planos distintos do \mathbb{R}^3 que contêm a origem, então $U + W = \mathbb{R}^3$ mas $\dim(U) + \dim(W) > 3$.

22. Sejam U_1 e U_2 subespaços de um espaço vetorial finitamente gerado V . Mostre que

$$\dim(U_1 + U_2) = \dim(U_1) + \dim(U_2) - \dim(U_1 \cap U_2).$$

SUGESTÃO: Seja B uma base de $U_1 \cap U_2$. Determine um conjunto S_1 de vetores que completa B para uma base de U_1 e um conjunto S_2 que completa B para uma base de U_2 . Mostre que $B \cup S_1 \cup S_2$ é uma base de V e conclua a equação desejada contanto os elementos de $B \cup S_1 \cup S_2$ de duas maneiras diferentes.

23. Sejam U e W subespaços vetoriais de \mathbb{R}^{10} de dimensões 8 e 9, respectivamente. Só há dois valores possíveis para a dimensão de $U \cap W$. Quais são estes valores?

24. Sejam U e W subespaços vetoriais de \mathbb{R}^n . Sabendo-se que U não está contido em W , nem W está contido em U , prove que existe um vetor de $U + W$ que não está em U nem em W .

25. Use o método de Gram-Schmidt para achar bases ortonormais para os subespaços dados abaixo:

- (a) $\langle(1, 2, 1), (2, 1, 1)\rangle \subset \mathbb{R}^3$;
 (b) $\langle(1, 1, 0, 1), (1, 0, -1, 1)\rangle \subset \mathbb{R}^4$;
 (c) $\langle(1, 1, 0, 1), (1, 0, -1, 1), (1, 1, 1, 0)\rangle \subset \mathbb{R}^4$.

26. Determine os complementos ortogonais de cada um dos espaços do exercício anterior.

27. Seja S um subespaço de \mathbb{R}^n . Calcule $\dim(S^\perp)$ em função de $\dim(S)$.

28. Determine o complemento ortogonal do subespaço de \mathbb{R}^3 definido por

$$\{(x, y, z) \in \mathbb{R}^3 \mid x + 2y + z = 0\}.$$

29. Determine o complemento ortogonal do subespaço do \mathbb{R}^3 gerado por $(1, 0, 0)$, $(0, 1, 1)$ e $(1, -1, -1)$.

CAPÍTULO 7

Transformações lineares

Aplicações são o instrumento matemático que usamos quando queremos efetuar transformações sobre um conjunto, ou quando desejamos comparar dois conjuntos. Como já vimos no caso do plano os operadores lineares proveem uma maneira adequada de fazer isto quando os objetos sobre os quais estamos operando são vetores. Neste capítulo consideraremos a noção mais geral de transformação linear e estudaremos suas propriedades básicas.

1. Definição e exemplos

Começamos generalizando o conceito de operador linear do plano para cobrir, não apenas operadores do \mathbb{R}^n , mas também aplicações entre espaços de dimensões diferentes.

1.1. Definição. Sejam V e W subespaços do \mathbb{R}^n e do \mathbb{R}^m , respectivamente. Generalizando o que aprendemos no caso do plano, diremos que uma *transformação linear* é uma função

$$T : V \rightarrow W$$

que satisfaz

$$(124) \quad T(u + v) = T(u) + T(v) \quad \text{e} \quad T(\lambda u) = \lambda T(u);$$

quaisquer que sejam $u, v \in V$ e $\lambda \in \mathbb{R}$. Reservaremos o nome *operador linear* para o caso em que $V = W$. Portanto, quando nos referirmos a uma função como um operador, estamos implicitamente supondo que se trata de uma transformação linear para a qual os conjuntos de partida e de chegada são os mesmos.

É importante observar, desde já, que a *imagem* de uma transformação linear $T : V \rightarrow W$, definida por

$$\text{Im}(T) = \{T(v) \mid v \in V\} \subset W,$$

não coincide necessariamente com o espaço de chegada W . Isto ocorre mesmo com operadores lineares do plano; por exemplo, a imagem de uma projeção do plano sobre uma reta é a própria reta, embora o conjunto de chegada seja todo o plano. Investigaremos as propriedades da imagem de uma transformação linear em detalhe na seção 2.

Observe que não dissemos como uma transformação linear age sobre o vetor nulo. A razão é que isto é consequência das duas propriedades enunciadas em (124). De fato, se v é um vetor qualquer de V , então

$$T(0) = T(0 \cdot u) = 0 \cdot T(u) = 0,$$

pois o escalar zero multiplicado por qualquer vetor dá zero. Ao longo deste capítulo estudaremos várias outras propriedades das transformações lineares que, como o fato de que levam o vetor nulo nele próprio, dependem apenas de (124) e de propriedades do \mathbb{R}^n , como a existência de bases para seus subespaços e o fato de que toda base de um subespaço pode ser completada para uma base do \mathbb{R}^n .

Transformações lineares de \mathbb{R}^n em \mathbb{R}^m sempre podem ser representadas por matrizes. Para ver porque isto é verdade, basta generalizar o que fizemos quando $m = n = 2$. Seja, portanto, $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ uma transformação linear. Se

$$(125) \quad v = (x_1, \dots, x_n) \in \mathbb{R}^n,$$

podemos escrevê-lo na forma

$$v = x_1e_1 + \cdots + x_ne_n,$$

em que $\varepsilon = \{e_1, \dots, e_n\}$ é a base canônica do \mathbb{R}^n . Aplicando T a v e utilizando as propriedades que definem uma transformação linear, segue-se que

$$T(v) = x_1T(e_1) + \cdots + x_mT(e_m).$$

Logo, se

$$T(e_j) = (a_{1j}, \dots, a_{nj})$$

são as coordenadas de $T(e_j)$, então, representando estes vetores como matrizes coluna, obtemos

$$T(v) = \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} x_1 + \cdots + \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} x_n,$$

que podemos reescrever na forma AX com

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \quad \text{e} \quad X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Nada disto é uma grande novidade porque os exemplos de operadores lineares do plano que encontramos no capítulo 1 foram, todos eles, descritos em forma matricial. Na próxima seção veremos que, em última análise, toda transformação linear pode ser expressa em termos de matrizes. Antes disto, porém, veremos como generalizar projeções e reflexões para espaços de dimensão maior que dois.

1.2. Coordenadas e isomorfismos. Suponhamos, como acima, que V é um subespaço do \mathbb{R}^n e que

$$\beta = \{v_1, \dots, v_r\}$$

é uma base de V . Como no caso dos vetores do plano, as *coordenadas* de um vetor $v \in V$ relativamente à base β são os números reais a_1, \dots, a_r para os quais

$$(126) \quad v = a_1 v_1 + \cdots + a_r v_r.$$

Uma vez que tenhamos fixado a ordenação dos vetores em β , podemos escrever estas coordenadas como um vetor

$$(v)_\beta = (a_1, \dots, a_r) \in \mathbb{R}^r.$$

Para que esta definição faça sentido, é essencial que haja apenas uma escolha possível de a 's para os quais a igualdade (126) é verificada. Isto é consequência do fato de que β é uma base de V . De fato, se

$$a_1 v_1 + \cdots + a_r v_r = b_1 v_1 + \cdots + b_r v_r,$$

então, subtraindo um lado do outro, obtemos

$$(a_1 - b_1)v_1 + \cdots + (a_r - b_r)v_r = 0.$$

Como os vetores de uma base são linearmente independentes, podemos concluir que

$$a_1 - b_1 = \cdots = a_r - b_r = 0,$$

o que mostra que há uma única escolha possível para os coeficientes na combinação linear (126).

 Para que pudéssemos representar um vetor em termos de suas coordenadas de maneira única precisamos fixar uma ordem para os vetores de uma base. Para ser consistente, deveríamos, de agora em diante, descrever a base de um subespaço como uma lista, porque a convenção que vigora em matemática é que a ordem em que os elementos de um conjunto são escritos é indiferente. Infelizmente, em álgebra linear, a base de um subespaço é sempre escrita entre chaves, como se fosse um conjunto, apesar da ordem dos seus elementos ser essencial quando precisamos determinar coordenadas de um vetor. Seguiremos esta convenção (inconsistente) neste livro porque você vai se deparar com ela em qualquer outro texto que faça uso de métodos de álgebra linear.

Antes de continuar, precisamos de um exemplo. Seja S o conjunto solução do sistema

$$\begin{cases} x_1 + x_2 - x_3 + x_4 - x_5 = 0 \\ x_2 - 3x_3 - x_5 = 0. \end{cases}$$

Resolvendo o sistema por substituição, obtemos

$$x_1 = -2x_3 - x_4 \quad \text{e} \quad x_2 = 3x_3 + x_5.$$

Logo, $v = (x_1, x_2, x_3, x_4, x_5) \in S$ se, e somente se,

$$(x_1, x_2, x_3, x_4, x_5) = (-2x_3 - x_4, 3x_3 + x_5, x_3, x_4, x_5),$$

que equivale a

$$(x_1, x_2, x_3, x_4, x_5) = x_3(-2, 3, 1, 0, 0) + x_4(-1, 0, 0, 1, 0) + x_5(0, 1, 0, 0, 1).$$

Logo,

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

é uma base de V e

$$(v)_\alpha = (x_3, x_4, x_5) \in \mathbb{R}^3.$$

Isso significa que cada vetor $v \in S$ admite duas representações diferentes. Como $S \subset \mathbb{R}^5$, podemos escrever v como uma 5-upla, da maneira usual. Por outro lado, se desejarmos esquecer o espaço ambiente e nos concentrarmos apenas no que ocorre em S , podemos representar v em termos de suas coordenadas relativamente a uma base de S . Neste caso o mesmo vetor v será descrito como uma tripla de números, já que $\dim(S) = 3$. Este comentário se aplica, com as mudanças óbvias, a vetores em qualquer subespaço do \mathbb{R}^n .

Podemos reformular o que fizemos acima em termos de transformações lineares. Supondo que V é um subespaço do \mathbb{R}^n e que β é uma base de V , podemos definir uma transformação linear

$$T_\beta : V \rightarrow \mathbb{R}^k$$

que associa a cada vetor $v \in V$ seu vetor de coordenadas $(v)_\beta \in \mathbb{R}^k$. A transformação T construída desta maneira está bem definida porque, como já vimos, a cada vetor de V está associada uma única k -upla de coordenadas na base β . Esta transformação é inversível. Para provar isto, considere a transformação linear

$$S_\beta : \mathbb{R}^k \rightarrow V$$

definida por

$$S_\beta(a_1, \dots, a_k) = a_1v_1 + \dots + a_kv_k.$$

Como

$$T_\beta(v) = (v)_\beta = (a_1, \dots, a_k),$$

temos, pela definição de coordenadas na base β que

$$S_\beta(T_\beta(v)) = S_\beta(a_1, \dots, a_k) = a_1v_1 + \dots + a_kv_k = v;$$

de modo que $S_\beta \circ T_\beta = \text{id}$. Por outro lado,

$$T_\beta(S_\beta(a_1, \dots, a_k)) = T_\beta(a_1v_1 + \dots + a_kv_k) = T(v) = (a_1, \dots, a_k);$$

onde $T_\beta \circ S_\beta = \text{id}$. Aplicando isto ao exemplo, verificamos que

$$S_\alpha : \mathbb{R}^3 \rightarrow V$$

é definida por

$$S_\alpha(x_3, x_4, x_5) = (-2x_3 - x_4, 3x_3 + x_5, x_3, x_4, x_5).$$

Quando uma transformação linear é inversível dizemos que é um *isomorfismo* e que o domínio e o contra-domínio desta transformação são espaços *isomórfos*. Indicamos que V

e W são isomorfos, escrevendo $V \cong W$. Assim, a transformação T_β construída acima é um isomorfismo entre V e \mathbb{R}^k ; de modo que $V \cong \mathbb{R}^k$.

A importância dos isomorfismos está no fato de que espaços isomorfos têm muitas propriedades em comum, entre elas a dimensão. Contudo, como os isomorfismos, em geral, não preservam o produto interno, as propriedades que dependem deste produto não são necessariamente preservadas.

1.3. Projeções. Ao contrário do que acontecia com o plano, em que só podíamos projetar um vetor sobre outro, no caso do \mathbb{R}^n , podemos projetá-lo sobre qualquer subespaço próprio. Além disso, podemos considerar a projeção como uma transformação que retorna vetores em um espaço de dimensão menor, o que será muito útil nas aplicações, como veremos no próximo artigo. Começaremos analisando a projeção sobre uma reta, porque é semelhante ao caso análogo do plano.

Seja, então, $u \in \mathbb{R}^n$ o vetor diretor unitário de uma reta. A projeção de um vetor v sobre u pertence à reta $\langle u \rangle$. Como u é unitário, basta determinar o comprimento da projeção de v sobre u , que é igual a $\|v\| \cos(\theta)$, em que θ é o ângulo entre u e v . Contudo, pela definição do ângulo entre dois vetores, temos que

$$\cos(\theta) = \frac{\langle v | u \rangle}{\|v\|},$$

pois $\|u\| = 1$ por hipótese. Assim, a projeção de v sobre u terá norma igual a

$$\|v\| \cos(\theta) = \langle v | u \rangle = u^t v.$$

Portanto, o vetor projeção desejado será igual a

$$(u^t v)u = u(u^t v) = (uu^t)v$$

pois, sendo um número, $u^t v$ comuta com u . Temos, assim, que a matriz que define a projeção sobre a reta $\langle u \rangle$ é igual a uu^t .

Para projetar um vetor qualquer $v \in \mathbb{R}^n$ sobre um subespaço U , precisamos de uma base ortonormal

$$\beta = \{u_1, \dots, u_k\}$$

de U . Projetamos, então, v sobre cada vetor de β e somamos os vetores resultantes para obter a projeção de v sobre U . Assim,

$$\text{Proj}_U(v) = (u_1 u_1^t + \dots + u_k u_k^t)v.$$

Portanto,

$$(127) \quad u_1 u_1^t + \dots + u_k u_k^t$$

nos dá a matriz desta projeção sobre a base canônica.

Como no caso do plano, a matriz (127) descreve um operador do \mathbb{R}^n cuja imagem é o subespaço U . Contudo, em muitas aplicações é mais conveniente trabalhar com a composta

$$P_U = T_\beta \circ \text{Proj}_U : \mathbb{R}^n \rightarrow \mathbb{R}^k$$

em que T_β é a transformação definida no artigo 1.2. Assim,

$$P_U(v) = T_\beta(\text{Proj}_U(v)) = (\text{Proj}_U(v))_\beta.$$

Como a base β é ortonormal, esta transformação é extremamente fácil de explicitar, porque temos de (127) que, se $v \in U$, então

$$\text{Proj}_U(v) = (u_1^t v)u_1 + \cdots + (u_k^t v)u_k.$$

Portanto,

$$P_U(v) = (\text{Proj}_U(v))_\beta = (u_1^t v, \dots, u_k^t v).$$

Podemos obter este último vetor, facilmente, fazendo o produto

$$\begin{bmatrix} - & u_1^t & - \\ \vdots & \vdots & \vdots \\ - & u_k^t & - \end{bmatrix} v = \begin{bmatrix} u_1^t v \\ \vdots \\ u_k^t v \end{bmatrix}.$$

Assim, a transformação linear $P_U : \mathbb{R}^n \rightarrow \mathbb{R}^k$ que estamos considerando é descrita pela matriz $k \times n$

$$\begin{bmatrix} - & u_1^t & - \\ \vdots & \vdots & \vdots \\ - & u_k^t & - \end{bmatrix}.$$

Aplicaremos o que acabamos de fazer para desenhar a projeção de um hipercubo quadrídimensional no espaço tridimensional. Como cada projeção do hipercubo envolve uma quantidade razoável de cálculos, nossa meta é descrever um algoritmo que nos permita facilmente desenhar a projeção desejada. Naturalmente, precisamos começar definindo o que é um hipercubo.

Nosso ponto de partida será o quadrado, que podemos considerar como uma versão plana do cubo. O quadrado unitário que queremos descrever é aquele que tem vértices na origem e nos pontos $(1, 0)$, $(0, 1)$ e $(1, 1)$. Seguindo nossa prática usual, diremos que este quadrado tem vértices nos vetores 0 , e_1 , e_2 e $e_1 + e_2$. Nem todos os vértices são ligados entre si por lados, mas somente aqueles que são adjacentes: para identificá-los basta notar que correspondem àqueles que diferem em apenas uma de suas coordenadas.

Passando, agora ao cubo cujos vértices são os vetores

$$0, e_1, e_2, e_3, e_1 + e_2, e_1 + e_3, e_2 + e_3 \quad \text{e} \quad e_1 + e_2 + e_3,$$

é fácil verificar que, mais uma vez, dois vértices são ligados por uma aresta se, e somente se, têm apenas uma coordenada diferente. Seguindo este padrão, definimos o hipercubo como sendo a figura do \mathbb{R}^4 que tem vértices em

$$(128) \quad 0, e_1, e_2, e_3, e_4, e_1 + e_2, e_1 + e_3, e_1 + e_4, e_2 + e_3, e_2 + e_4, e_3 + e_4, e_1 + e_2 + e_3, \\ e_1 + e_2 + e_4, e_1 + e_3 + e_4, e_2 + e_3 + e_4 \quad \text{e} \quad e_1 + e_2 + e_3 + e_4.$$

Portanto, assim, como um quadrado tem 4 vértices e um cubo tem 8 vértices, um hipercubo tem 16 vértices. Como no caso das outras duas figuras, somente os vértices do hipercubo que diferem em uma única coordenada são ligados por arestas.

Tendo definido o hipercubo, podemos projetá-lo no \mathbb{R}^3 usando a transformação P_U cuja matriz é

$$\begin{bmatrix} - & u_1^t & - \\ - & u_2^t & - \\ - & u_3^t & - \end{bmatrix}.$$

Partiremos do princípio de que a escolha do subespaço de dimensão três sobre o qual será feita a projeção fica à escolha de quem for aplicar o algoritmo. A melhor maneira de fazer isto é permitir que a entrada do algoritmo sejam três vetores linearmente independentes do \mathbb{R}^4 . Contudo, como a base sobre a qual a projeção é feita deve ser ortonormal, será necessário aplicar o algoritmo de Gram-Schmidt a esses vetores antes de construir a matriz de P_U .

Por exemplo, se escolhermos o subespaço gerado por

$$(1, 1, 0, 0), (0, 0, 1, 1) \quad \text{e} \quad (0, 1, 0, 1)$$

o algoritmo de Gram-Schmidt retornará os vetores

$$u_1 = \frac{1}{\sqrt{2}}(1, 1, 0, 0), \quad u_2 = \frac{1}{\sqrt{2}}(0, 0, 1, 1) \quad \text{e} \quad u_3 = \frac{1}{2}(1, -1, 1, -1),$$

de modo que a matriz que dá a projeção será

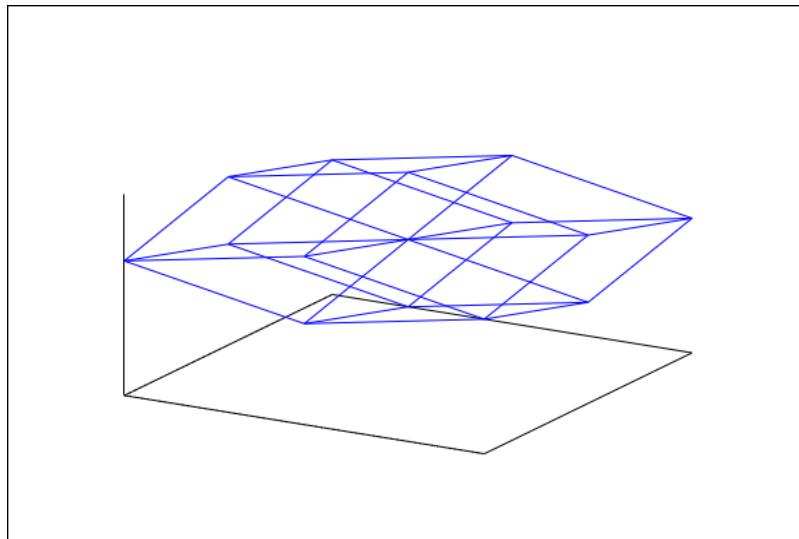
$$\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix}$$

As projeções dos vários vértices são dadas pela tabela 1 e a figura resultante da projeção pode ser vista na figura 1.

1.4. Reflexões. Observe, primeiramente, que há uma maneira alternativa de definir a projeção quando o subespaço é um hiperplano. Sejam $v \in \mathbb{R}^n$ e H um hiperplano do \mathbb{R}^n

Vértices	Projeções	Vértices	Projeções
0	(0, 0, 0)	$e_2 + e_3$	$(1/\sqrt{2}, 1/\sqrt{2}, 0)$
e_1	$(1/\sqrt{2}, 0, 1/2)$	$e_2 + e_4$	$(1/\sqrt{2}, 1/\sqrt{2}, -1)$
e_2	$(1/\sqrt{2}, 0, -1/2)$	$e_3 + e_4$	$(0, \sqrt{2}, 0)$
e_3	$(0, 1/\sqrt{2}, 1/2)$	$e_1 + e_2 + e_3$	$(\sqrt{2}, 1/\sqrt{2}, 1/2)$
e_4	$(0, 1/\sqrt{2}, -1/2)$	$e_1 + e_2 + e_4$	$(\sqrt{2}, 1/\sqrt{2}, -1/2)$
$e_1 + e_2$	$(\sqrt{2}, 0, 0)$	$e_1 + e_3 + e_4$	$(1/\sqrt{2}, \sqrt{2}, 1/2)$
$e_1 + e_3$	$(1/\sqrt{2}, 1/\sqrt{2}, 1)$	$e_2 + e_3 + e_4$	$(1/\sqrt{2}, \sqrt{2}, -1/2)$
$e_1 + e_4$	$(1/\sqrt{2}, 1/\sqrt{2}, 0)$	$e_1 + e_2 + e_3 + e_4$	$(\sqrt{2}, \sqrt{2}, 0)$

TABELA 1. Projeções dos vértices do hipercubo

FIGURA 1. Projeção do hipercubo no subespaço U

cujo vetor *normal unitário* denotaremos por u . Subtraindo de v sua componente ao longo de u , obtemos sua projeção sobre H , porque

$$\langle v - \langle u | v \rangle u | u \rangle = \langle v | u \rangle - \langle u | v \rangle \langle u | u \rangle = 0.$$

Escrevendo esta fórmula para a projeção em termos de matrizes, temos

$$v - \langle u | v \rangle u = v - uu^t v = (I - uu^t)v,$$

de modo que a matriz da projeção sobre o hiperplano H pode ser escrita na forma

$$I - uu^t.$$

Por exemplo, o plano de equação $x + y + z = 0$ tem

$$u = \frac{1}{\sqrt{3}}(1, 1, 1);$$

por vetor normal unitário. Logo, a matriz que realiza a projeção do \mathbb{R}^3 neste plano é

$$I - u^t u = I - \frac{1}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$$

que é igual a

$$\frac{1}{3} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}.$$

Esta maneira de projetar nos permite facilmente refletir um vetor v do \mathbb{R}^n relativamente ao hiperplano H , que faz o papel de espelho. Seja u um vetor unitário, ortogonal a H . Se subtrair $uu^t \cdot v$ de v produz um vetor em H , então subtraindo $uu^t \cdot v$ duas vezes obtemos um vetor que satisfaz

$$\text{Proj}_H(v - 2uu^t \cdot v) = \text{Proj}_H(v) \quad \text{e} \quad \text{Proj}_u(v - 2uu^t \cdot v) = -\text{Proj}_u(v).$$

Mas essas são as propriedades que definem a reflexão, tanto no \mathbb{R}^2 , quanto no \mathbb{R}^3 . Por isso chamamos de *reflexão* o operador do \mathbb{R}^n que satisfaz estas mesmas propriedades. Como

$$v - 2uu^t \cdot v = (I - 2uu^t)v,$$

em que I é a matriz identidade $n \times n$, concluímos que a matriz da reflexão é

$$(129) \quad R_H = I - 2uu^t.$$

Por exemplo, para determinar a matriz da reflexão cujo espelho é o hiperplano $x + y + z + w = 0$, calculamos um vetor unitário perpendicular a este plano, digamos

$$u = \frac{1}{2}(1, 1, 1, 1).$$

Substituindo o vetor na fórmula (129), obtemos

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

Observe que, para achar a matriz de uma reflexão precisamos apenas conhecer um vetor unitário ortogonal ao espelho. Isto implica que uma reflexão R do \mathbb{R}^n fica completamente determinada se conhecemos um vetor v , fora do hiperplano de reflexão e sua imagem por

v , porque $v - R(v)$ tem que ser um vetor ortogonal ao espelho. Ainda que um desenho baste para lhe convencer de que isto vale no plano, ou até mesmo no espaço tridimensional, a única maneira de garantir que também é válido para o \mathbb{R}^n é usar a definição de reflexão ou a matriz que derivamos a partir desta definição.

Suponha, então, que $R = I - 2uu^t$ é uma reflexão do \mathbb{R}^n cujo espelho é o hiperplano H ortogonal ao vetor unitário u . Se $v \notin H$, então

$$v - R(v) = v - (v - uu^t v) = u(u^t v).$$

Como

$$u^t v = \langle u | v \rangle$$

é um escalar, mostramos que $v - R(v)$ é um múltiplo de u . Portanto, para achar $\pm u$ e calcular a matriz da reflexão basta normalizar o vetor $v - R(v)$.

Um exemplo muito importante desta última construção ocorre quando a reflexão R leva um dado vetor v de \mathbb{R}^n em um vetor colinear ao vetor e_1 da base canônica. Como o comprimento de um vetor não pode ser alterado por reflexão, deveremos ter que

$$R(v) = \|v\|e_1.$$

Portanto, o vetor unitário u normal ao espelho será obtido dividindo

$$v - \|v\|e_1$$

por seu comprimento. Diremos, neste caso, que R é a *reflexão de Householder* determinada por v . O nome é uma homenagem a A. Householder que mostrou em 1958 [4] como estas matrizes poderiam ser usadas em um algoritmo de decomposição matricial.

Por exemplo, quando $v = (1, 1, 1, 1) \in \mathbb{R}^4$, o vetor n é obtido normalizando-se

$$(1, 1, 1, 1) - 2(1, 0, 0, 0) = (-1, 1, 1, 1),$$

de modo que a reflexão de Householder correspondente tem matriz

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

As reflexões de Householder são usadas em um algoritmo para decomposição QR que tem várias propriedades vantagens sobre o algoritmo de Gram-Schmidt.

2. Núcleo e imagem

Nesta seção introduzimos o conceito de núcleo de uma transformação linear e investigamos as propriedades do núcleo e da imagem dessas transformações.

2.1. Núcleo e imagem de uma transformação linear. Como vimos no artigo 1.1, a *imagem* de uma transformação linear $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ é o conjunto dos vetores do \mathbb{R}^m que são imagem por T de algum vetor de \mathbb{R}^n ; assim,

$$\text{Im}(T) = \{Tv \mid v \in \mathbb{R}^n\}.$$

Nossa primeira observação é que $\text{Im}(T)$ é mais que um simples conjunto, ela é um subespaço de W . Isto é fácil de verificar, porque, para começar, $T(0) = 0$ implica que $0 \in \text{Im}(T)$, de modo que $\text{Im}(T)$ não pode ser vazio. Além disso, se $Tv, Tv' \in \text{Im}(T)$, então

$$Tv + Tv' = T(v + v') \in \text{Im}(T);$$

ao passo que, se $\lambda \in \mathbb{R}$, então temos também que

$$\lambda T(v) = T(\lambda v) \in \text{Im}(T).$$

Portanto, valem todas as propriedades necessárias para garantir que $\text{Im}(T)$ é um subespaço do \mathbb{R}^m .

Seja $\beta = \{u_1, \dots, u_n\}$ uma base qualquer do \mathbb{R}^n . Se $v \in \mathbb{R}^n$, então existem números reais a_1, \dots, a_n tais que

$$v = a_1u_1 + \cdots + a_nu_n.$$

Logo,

$$T(v) = a_1T(u_1) + \cdots + a_nT(u_n),$$

pela linearidade de T . Portanto,

$$(130) \quad \text{Im}(T) = \langle T(u_1), \dots, T(u_n) \rangle.$$

Diante disto, é inevitável não nos perguntarmos se estes vetores sempre formam uma base de $\text{Im}(T)$. A resposta é não. Para identificar onde está o problema, suponhamos que

$$a_1T(u_1) + \cdots + a_nT(u_n) = 0$$

é uma combinação linear destes vetores que é igual ao vetor nulo. Usando a linearidade de T , podemos reescrever isto na forma

$$T(a_1u_1 + \cdots + a_nu_n) = 0.$$

Se pudéssemos deduzir disto que

$$a_1u_1 + \cdots + a_nu_n = 0,$$

bastaria apelar para o fato de β ser base do \mathbb{R}^n para que pudéssemos concluir que todos os a 's têm que ser nulos. Portanto, se o obstáculo para que

$$T(u_1), \dots, T(u_n)$$

seja uma base da imagem de T está na existência de vetores não nulos no \mathbb{R}^n que T leva no vetor nulo do \mathbb{R}^m . Transformações para as quais tais vetores existem incluem todas as projeções do \mathbb{R}^n em subespaços próprios.

Isto põe em evidência a importância de identificar aqueles vetores que são levados no zero por uma transformação linear. O conjunto de todos estes vetores é o núcleo da transformação. Mais precisamente, o *núcleo* de uma transformação linear $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ é o conjunto

$$N(T) = \{v \in \mathbb{R}^n \mid Tv = 0\}.$$

Representando a transformação T por uma matriz A , como fizemos no artigo 1.1, vemos que o núcleo coincide com o conjunto solução do sistema $AX = 0$. Portanto, assim como no caso da imagem, o núcleo de uma transformação linear também é um subespaço. Note contudo, que o núcleo é um subespaço do espaço de partida, ao passo que a imagem é um subespaço do espaço de chegada, como ilustrado no diagrama

$$\begin{array}{ccc} \mathbb{R}^n & \xrightarrow{T} & \mathbb{R}^m \\ \uparrow N(T) & & \uparrow \text{Im}(T) \end{array}$$

em que as setas verticais indicam inclusão.

Por exemplo, suponhamos que $T : \mathbb{R}^4 \rightarrow \mathbb{R}^4$ seja a transformação linear definida por

$$T(x, y, z, w) = (x + y, y - z, z - w, y - w).$$

De acordo com (130), $\text{Im}(T)$ é gerada pela imagem dos vetores da base canônica do \mathbb{R}^4 por T , donde

$$\text{Im}(T) = \langle (1, 0, 0, 0), (1, 1, 0, 1), (0, -1, 1, 0), (0, 0, -1, -1) \rangle.$$

Aplicando eliminação gaussiana à matriz

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

obtemos

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

de modo que

$$\{(1, 0, 0, 0), (0, 1, 0, 1), (0, 0, 1, 1)\}$$

é uma base da imagem de T . Em particular $\text{Im}(T)$ tem dimensão 3. Note também que, neste exemplo, as imagens dos vetores da base canônica não são linearmente independentes. Pelas considerações anteriores, podemos concluir que isto indica que o núcleo de T

não deve ser nulo. Mas $N(T)$ é igual ao conjunto solução do sistema

$$\begin{aligned}x + y &= 0 \\y - z &= 0 \\z - w &= 0 \\y - w &= 0\end{aligned}$$

de modo que $(x, y, z, w) \in N(T)$ se, e somente se,

$$(x, y, z, w) = (-w, w, w, w) = w(-1, 1, 1, 1).$$

Logo o núcleo tem base

$$B = \{(-1, 1, 1, 1)\}$$

e dimensão um. Observe que, neste exemplo,

$$\dim(N(T)) + \dim(\text{Im}(T)) = 1 + 3 = 4$$

é igual a dimensão do conjunto de partida. Longe de ser mera coincidência, esta é uma importante propriedade que relaciona o núcleo e a imagem de uma transformação linear, como veremos no próximo artigo.

2.2. Teorema do núcleo e da imagem. Seja $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ uma transformação linear. Como vimos no artigo 2.1, podemos associar a T dois subespaços vetoriais: o núcleo, que é um subespaço do \mathbb{R}^n , e a imagem, que é um subespaço do \mathbb{R}^m . Ao longo deste artigo denotaremos por $T(X)$ o conjunto de todos os vetores obtidos aplicando T aos vetores de um subconjunto $X \subset \mathbb{R}^n$. Assim, se

$$X = \{v_1, \dots, v_k\} \subset \mathbb{R}^n,$$

então

$$T(X) = \{T(v_1), \dots, T(v_k)\} \subset \text{Im}(T) \subset \mathbb{R}^m.$$

Como vimos no artigo anterior, as imagens dos vetores em uma base do \mathbb{R}^n por uma transformação linear

$$T : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

sempre constituem um conjunto de geradores de $\text{Im}(T)$; mas este conjunto não é necessariamente uma base. Contudo, é possível construir uma base do \mathbb{R}^n a partir da qual obtemos, sem nenhum esforço extra, uma base da imagem de T .

Para construir uma tal base, começamos escolhendo uma base α do núcleo de T . Segundo o teorema 4.3, podemos acrescentar vetores linearmente independentes a β até obter uma base β do \mathbb{R}^n . Logo,

$$\beta = \alpha \cup \alpha',$$

em que α' é o conjunto dos vetores que precisamos acrescentar a α para obtermos uma base do \mathbb{R}^n . Portanto,

$$T(\beta) = T(\alpha) \cup T(\alpha'),$$

é um conjunto de geradores para $\text{Im}(T)$. Contudo, $T(\alpha) = \{0\}$, porque os vetores de α pertencem ao núcleo de T , de modo que

$$T(\beta) = T(\alpha') \cup \{0\}.$$

Isto significa que $T(\alpha')$ é um conjunto de geradores para $\text{Im}(T)$, já que o vetor nulo não contribui nada novo. Nossa meta é provar que $T(\alpha')$ é um conjunto de vetores linearmente independentes e, portanto, uma base para $\text{Im}(T)$.

Seja, então,

$$\alpha' = \{w_1, \dots, w_m\},$$

de modo que

$$T(\alpha') = \{T(w_1), \dots, T(w_m)\}$$

e suponhamos que

$$a_1T(w_1) + \dots + a_mT(w_m) = 0$$

é uma combinação linear nula dos vetores de $T(\alpha')$. Pela linearidade de T , obtemos

$$T(a_1w_1 + \dots + a_mw_m) = 0,$$

de modo que $a_1w_1 + \dots + a_mw_m \in N(T)$. Denotando por W o subespaço do \mathbb{R}^n gerado pelos elementos de W , podemos afirmar que

$$a_1w_1 + \dots + a_mw_m \in N(T) \cap W = \{0\},$$

pois W é um espaço complementar de $N(T)$. Isto implica que

$$a_1w_1 + \dots + a_mw_m = 0$$

e como os w s são linearmente independentes, segue-se que os a s são todos nulos, como queríamos mostrar. Portanto,

se α' for o conjunto de vetores acrescentados a uma base α do núcleo de T de modo a obter uma base do \mathbb{R}^n , então $T(\alpha')$ é uma base da imagem de T .

Por exemplo, o núcleo da transformação $T : \mathbb{R}^4 \rightarrow \mathbb{R}^4$ definida por $T(x, y, z, w) = (x + y, 0, 0, z)$ é o conjunto dos vetores (x, y, z, w) para os quais

$$T(x, y, z, w) = (x + y, 0, 0, z) = (0, 0, 0, 0);$$

que corresponde a ter $y = -x$ e $z = 0$. Isto é,

$$N(T) = \{(x, -x, 0, w) \mid x, w \in \mathbb{R}\}.$$

Como

$$(x, -x, 0, w) = x(1, -1, 0, 0) + w(0, 0, 0, 1),$$

e os vetores $(1, -1, 0, 0)$ e $(0, 0, 0, 1)$ são linearmente independentes, então

$$\beta = \{(1, -1, 0, 0), (0, 0, 0, 1)\}$$

é uma base de $N(T)$. Para completá-la, basta acrescentar dois vetores linearmente independentes a estes dois. Digamos que escolhemos os vetores $(0, 0, 1, 0)$ e $(1, 0, 0, 0)$; neste caso,

$$\beta' = \{(0, 0, 1, 0), (1, 0, 0, 0)\}.$$

Portanto, o que provamos acima nos garante que

$$T(\beta') = \{(0, 0, 0, 1), (1, 0, 0, 0)\}$$

é uma base de $\text{Im}(T)$, como é fácil de verificar diretamente.

Resumindo, mostramos que se $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ é uma transformação linear e se

- α é uma base de $N(T)$;
- α' é uma base de um complementar de $N(T)$;

então $T(\alpha')$ é uma base de $\text{Im}(T)$. Em particular,

$$\dim(N(T)) + \dim(\text{Im}(T)) = |\alpha| + |\alpha'| = |\alpha \cup \alpha'|,$$

em que a última igualdade vem do fato de que α e α' não têm elementos em comum. Como $\beta = \alpha \cup \alpha'$ é uma base de V ,

$$|\alpha \cup \alpha'| = \dim(V).$$

Combinando tudo isto, provamos o seguinte resultado.

TEOREMA DO NÚCLEO E DA IMAGEM. *Se $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ é uma transformação linear, então*

$$\dim(V) = \dim(N(T)) + \dim(\text{Im}(T)).$$

Uma aplicação simples deste teorema está relacionada à descrição de quando uma transformação linear

$$T : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

é injetiva, sobrejetiva ou bijetiva. Lembre-se que T é sobrejetiva quando

$$\text{Im}(T) = \mathbb{R}^m,$$

de maneira que a imagem coincide com o conjunto de chegada; ao passo que T é injetiva se vetores diferentes do \mathbb{R}^n são sempre levados por T em vetores diferentes do \mathbb{R}^m . Finalmente, T é bijetiva quando é simultaneamente sobrejetiva e injetiva.

O fato de T ser linear nos dá um critério bastante conveniente para determinar se é ou não injetiva, porque se $v, v' \in \mathbb{R}^n$ têm a mesma imagem, então

$$T(v - v') = T(v') - T(v') = 0;$$

de modo que $v - v' \in N(T)$. Como a recíproca é claramente verdadeira, podemos afirmar que *uma transformação linear é injetiva se, e somente se, seu núcleo é nulo*. Combinando isto com o teorema do núcleo e da imagem, temos que T só pode ser injetiva quando

$$n = \dim(N(T)) + \dim(\text{Im}(T)) = \dim(\text{Im}(T)) \leq m,$$

em que a desigualdade é consequência do corolário 4.4 aplicado à inclusão $\text{Im}(T) \subset \mathbb{R}^m$.

Por outro lado, se T é sobrejetora, então

$$n = \dim(N(t)) + \dim(\text{Im}(T)) = \dim(N(t)) + m;$$

onde

$$\dim(N(T)) = n - m.$$

Portanto, sempre que $n > m$ a transformação linear terá núcleo não nulo e não poderá ser injetiva. Combinando tudo isto concluímos que T só pode ser bijetiva quando $m = n$. A importância desta análise está no fato de que uma função é inversível se, e somente se, for bijetiva. Aplicando isto às transformações lineares, concluímos que uma transformação linear só pode ser um isomorfismo se $m = n$ e se tem núcleo igual ao espaço nulo.

3. Mudança de base

Nesta seção veremos como e porque é importante podermos descrever as matrizes de uma transformação linear relativamente a outras bases que não a canônica. Além disso, estenderemos a bases quaisquer e espaços de qualquer dimensão o que aprendemos sobre mudança de base no capítulo 2.

3.1. Matrizes e bases. Começamos generalizando a construção da matriz de uma transformação linear para o caso em que as bases consideradas no domínio e contradomínio não são as bases canônicas desses dois espaços. Seja $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ uma transformação linear. Se

$$\beta = \{u_1, \dots, u_n\}$$

é uma base do \mathbb{R}^n , podemos escrever $v \in \mathbb{R}^n$ na forma

$$(131) \quad v = b_1 u_1 + \dots + b_n u_n,$$

em que b_1, \dots, b_n são números reais. Aplicando T a esta expressão, obtemos

$$(132) \quad T(v) = b_1 T(u_1) + \dots + b_n T(u_n).$$

Supondo que

$$\beta' = \{w_1, \dots, w_m\}$$

é uma base do \mathbb{R}^m , teremos

$$T(u_j) = a_{1j} w_1 + \dots + a_{mj} w_m;$$

ou, o que é equivalente,

$$(T(u_j))_{\beta'} = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix},$$

para $1 \leq j \leq n$. Logo,

$$(T(v))_{\beta'} = b_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + \cdots + b_n \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix};$$

onde

$$(T(v))_{\beta'} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix},$$

que podemos escrever, de maneira mais compacta, como

$$(133) \quad (Tv)_{\beta'} = (T)_{\beta\beta'}(v)_{\beta},$$

em que

$$(134) \quad (T)_{\beta\beta'} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix} = \begin{bmatrix} | & & | \\ (Tu_1)_{\beta'} & \cdots & (Tu_n)_{\beta'} \\ | & & | \end{bmatrix}$$

é a matriz cujas colunas contêm as coordenadas dos vetores $T(u_1), \dots, T(u_n)$ na base β' .

Digamos, por exemplo, que queremos calcular $(T)_{\beta\beta'}$ quando

$$T : \mathbb{R}^2 \rightarrow \mathbb{R}^4$$

é a transformação linear definida por

$$(135) \quad T(x_1, x_2) = (2x_1 - x_2, 3x_1 + 9x_2, x_1 - x_2, -2x_1 + x_2)$$

e escolhemos

$$\beta = \{(1, 1), (1, 0)\}$$

para base do \mathbb{R}^2 e

$$\beta' = \left\{ \frac{1}{\sqrt{2}}(1, 1, 0, 0), \frac{1}{\sqrt{2}}(1, -1, 0, 0), \frac{1}{\sqrt{2}}(0, 0, 1, 1), \frac{1}{\sqrt{2}}(0, 0, 1, -1) \right\}$$

para a base do \mathbb{R}^4 . Como nada é dito sobre a base relativamente à qual a fórmula (135) está expressa, devemos supor que se trata da base canônica. Logo,

$$T(1, 1) = (1, 12, 0, -1) \quad \text{e} \quad T(1, 0) = (2, 3, 1, -2)$$

são as coordenadas destes dois vetores relativamente à base canônica do \mathbb{R}^4 . Como β' é uma base ortonormal do \mathbb{R}^4 , basta calcular produtos internos para obter as coordenadas destes mesmos vetores relativamente a β' . Fazendo isto, encontramos

$$(T(1, 1))_{\beta'} = \frac{1}{\sqrt{2}}(13, -11, -1, 1) \quad \text{e} \quad (T(1, 0))_{\beta'} = \frac{1}{\sqrt{2}}(5, -1, -1, 3).$$

Portanto,

$$(T)_{\beta\beta'} = \frac{1}{\sqrt{2}} \begin{bmatrix} 13 & 5 \\ -11 & -1 \\ -1 & -1 \\ 1 & 3 \end{bmatrix}.$$

Além do caso em que β é a base canônica do \mathbb{R}^n e β' a base canônica do \mathbb{R}^m , a fórmula (133) e a matriz (134) já haviam sido consideradas anteriormente quando $n = m = 2$. Neste último caso, $(T)_{\beta'\beta'}$ coincide com a matriz $(T)_{\beta'}$ definida no artigo 3.1 do capítulo 1. Como veremos no próximo artigo, as fórmulas (133) e (134) também nos permitem chegar, facilmente, às fórmulas de mudança de base para vetores e operadores lineares.

3.2. Mudança de base. Sejam β e β' duas bases de \mathbb{R}^n . Nosso objetivo neste artigo é relacionar as coordenadas de um vetor v relativas à β às suas coordenadas relativas a β' . O que fizemos no artigo anterior torna isto bastante fácil; basta aplicar as fórmulas (133) e a matriz (134) ao operador identidade do \mathbb{R}^n . Fazendo isto, as duas fórmulas nos dão,

$$(136) \quad (v)_{\beta'} = (\text{id})_{\beta\beta'}(v)_{\beta}, \quad \text{em que} \quad (\text{id})_{\beta\beta'} = \begin{bmatrix} | & & | \\ (u_1)_{\beta'} & \cdots & (u_n)_{\beta'} \\ | & & | \end{bmatrix}.$$

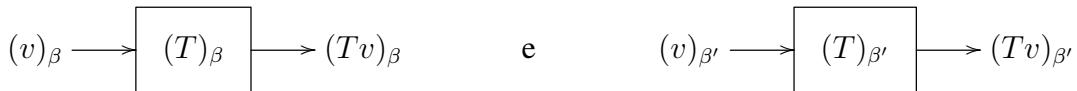
Quando $\beta' = \varepsilon$ é a base canônica do \mathbb{R}^n , as colunas desta matriz corresponderão aos vetores de β escritos em suas coordenadas relativamente a ε . Isso explica o porquê da notação introduzida no artigo 3.2 do capítulo 2 para a matriz de mudança de base no \mathbb{R}^2 .

Em vez de deduzir uma fórmula geral para a mudança de base de uma transformação linear entre dois espaços diferentes, vamos nos concentrar no caso que aparece com mais frequência nas aplicações, que é aquele em que T é um operador do \mathbb{R}^n . Neste caso, para simplificar a notação, escreveremos $(T)_{\beta}$ em vez de $(T)_{\beta\beta}$. Como no caso do plano, nossa meta é descobrir como transformar $(T)_{\beta}$ em $(T)_{\beta'}$ usando a matriz de mudança de base. Para isso, procedemos exatamente como fizemos no caso do plano.

Se $v \in \mathbb{R}^n$, então (133) nos dá

$$(137) \quad (Tv)_{\beta} = (T)_{\beta}(v)_{\beta} \quad \text{e} \quad (Tv)_{\beta'} = (T)_{\beta'}(v)_{\beta'}.$$

Como no caso do plano, podemos ilustrar estas fórmulas usando os diagramas:



Já a fórmula de mudança de base

$$(v)_{\beta'} = (\text{id})_{\beta\beta'}(v)_{\beta}$$

corresponde ao diagrama

$$(v)_\beta \longrightarrow \boxed{(\text{id})_{\beta\beta'}} \longrightarrow (v)_{\beta'}$$

Portanto,

$$(v)_\beta \longrightarrow \boxed{(\text{id})_{\beta\beta'}} \xrightarrow{(v)_{\beta'}} \boxed{(T)_{\beta'}} \xrightarrow{(Tv)_{\beta'}} \boxed{(\text{id})_{\beta'\beta}} \longrightarrow (Tv)_\beta$$

comporta-se, exatamente, como

$$(v)_\beta \longrightarrow \boxed{(T)_\beta} \longrightarrow (Tv)_\beta$$

o que nos leva a esperar que codifiquem as mesmas matrizes; isto é, que

$$(138) \quad (T)_\beta = (\text{id})_{\beta'\beta}(T)_{\beta'}(\text{id})_{\beta\beta'},$$

sem esquecer que é necessário inverter as posições das matrizes quando passamos dos diagramas às fórmulas e vice-versa.

Por mais convincente que seja o argumento que utiliza os diagramas, não se trata de uma demonstração. Para obtê-la, precisamos reproduzir o que foi feito com os diagramas usando a fórmula (137). Multiplicando a matriz do lado direito de (138) à direita pela matriz coluna das coordenadas de um vetor $v \in \mathbb{R}^n$ relativamente à base β , obtemos

$$(\text{id})_{\beta'\beta}(T)_{\beta'}(\text{id})_{\beta\beta'}(v)_\beta = (\text{id})_{\beta'\beta}(T)_{\beta'}(v)_{\beta'}$$

pela fórmula de mudança de base. Segue-se, então, de (137) que

$$(\text{id})_{\beta'\beta}(T)_{\beta'}(\text{id})_{\beta\beta'}(v)_\beta = (\text{id})_{\beta'\beta}(Tv)_{\beta'}.$$

Mas, aplicando novamente a fórmula de mudança de base, desta vez com a ordem das bases trocada, vemos que

$$(\text{id})_{\beta'\beta}(T)_{\beta'}(\text{id})_{\beta\beta'}(v)_\beta = (Tv)_\beta.$$

Assim, pela segunda fórmula em (137),

$$(\text{id})_{\beta'\beta}(T)_{\beta'}(\text{id})_{\beta\beta'}(v)_\beta = (T)_\beta(v)_\beta.$$

Como esta igualdade vale, quaisquer que sejam os vetores $v \in \mathbb{R}^n$, podemos concluir que (138) é verdadeira, como queríamos mostrar.

Por outro lado, pela fórmula de mudança de base,

$$(\text{id})_{\beta\beta'}(\text{id})_{\beta'\beta}(v)_{\beta'} = (\text{id})_{\beta\beta'}(v)_\beta = (v)_{\beta'}.$$

Levando em conta que esta igualdade também vale quaisquer que sejam os vetores $v \in \mathbb{R}^n$, obtemos

$$(\text{id})_{\beta\beta'}(\text{id})_{\beta'\beta} = I,$$

em que I é a matriz identidade $n \times n$. Portanto,

$$(\text{id})_{\beta\beta'} = (\text{id})_{\beta'\beta}^{-1},$$

o que nos permite reescrever (138) na forma

$$(139) \quad (T)_\beta = (\text{id})_{\beta'\beta}(T)_{\beta'}(\text{id})_{\beta'\beta}^{-1}.$$

Na maioria das aplicações que faremos pelo resto do livro, a mudança de base envolverá a base canônica e alguma outra base β do \mathbb{R}^n . Como, nestes casos, a matriz mais fácil de obter costuma ser $(\text{id})_{\beta\varepsilon}$, é conveniente enunciar a fórmula acima neste caso particular para futura referência:

$$(140) \quad (T)_\varepsilon = (\text{id})_{\beta\varepsilon}(T)_\beta(\text{id})_{\beta\varepsilon}^{-1};$$

sem esquecer que $(\text{id})_{\beta\varepsilon}$ é obtida escrevendo-se nas colunas as coordenadas dos vetores de β na base canônica. Os demais artigos desta seção serão dedicados a várias aplicações desta fórmula.

3.3. Rotações no \mathbb{R}^3 . Ao passo que todos temos uma compreensão bastante clara do que deveria ser uma rotação do plano, o mesmo não pode ser dito sobre as rotações do \mathbb{R}^3 . Em 1776, L. Euler publicou um artigo em que define rotação como qualquer transformação que pode ser aplicada a uma esfera cujo centro está fixo. A realização física mais próxima disto encontra-se nas fontes em forma de uma esfera de pedra extremamente polida que se encaixa perfeitamente em um suporte de mesma curvatura. Jatos de água entre a esfera e o suporte geram uma junta hidrostática, quase sem atrito, que permite a uma pessoa girar a esfera com bastante facilidade. A esfera da figura 2, situada no centro de visitantes de Cabo Canaveral, na Flórida, tem 9 toneladas, mas pode ser movida, sem dificuldade, por uma única pessoa.



FIGURA 2. Esfera de Cabo Canaveral

Como a esfera da fonte não está fixa em nenhum ponto, é possível rodá-la em qualquer direção, mas não movê-la para fora do suporte, fazendo com que seu centro permaneça

sempre no mesmo lugar apesar de não estar preso a nada. Usando esta definição Euler provou que toda rotação do \mathbb{R}^3 tem um eixo; isto é, uma reta que permanece fixa e em torno da qual a esfera gira. Isto significa, por exemplo, que se rodarmos uma esfera relativamente a dois eixos, uma rotação imediatamente seguida da outra, obteremos uma nova rotação relativa a um terceiro eixo, diferente dos dois com que começamos.

Um estudo mais detalhado das rotações, incluindo a demonstração do teorema de Euler, ficará para a seção 3 do próximo capítulo, porque requer ideias relacionadas a autovalores e autovetores. Neste artigo veremos apenas como usar mudança de base para construir a matriz na base canônica de uma rotação ρ cujo eixo e ângulo conhecemos. Porém, há dois comentários importantes que precisamos fazer sobre o ângulo de rotação, antes de podermos começar a construção da matriz. O primeiro é que *ele é sempre medido sobre o plano ortogonal ao eixo*. Isto é essencial, porque, como ilustrado na figura 3, o ângulo entre os vetores v e $\rho(v)$ decresce à medida que v se aproxima dos pólos. Usar um vetor ortogonal ao eixo garante que todos estamos falando sobre o mesmo ângulo. O segundo comentário diz respeito ao sentido da rotação. Neste artigo, vamos nos contentar em produzir a matriz de uma rotação de eixo e ângulo especificados, sem nos preocupar com seu sentido. Também voltaremos a esta questão, em mais detalhe, quando estudarmos rotações no próximo capítulo; lá veremos como especificar o sentido da rotação usando a regra da mão direita.

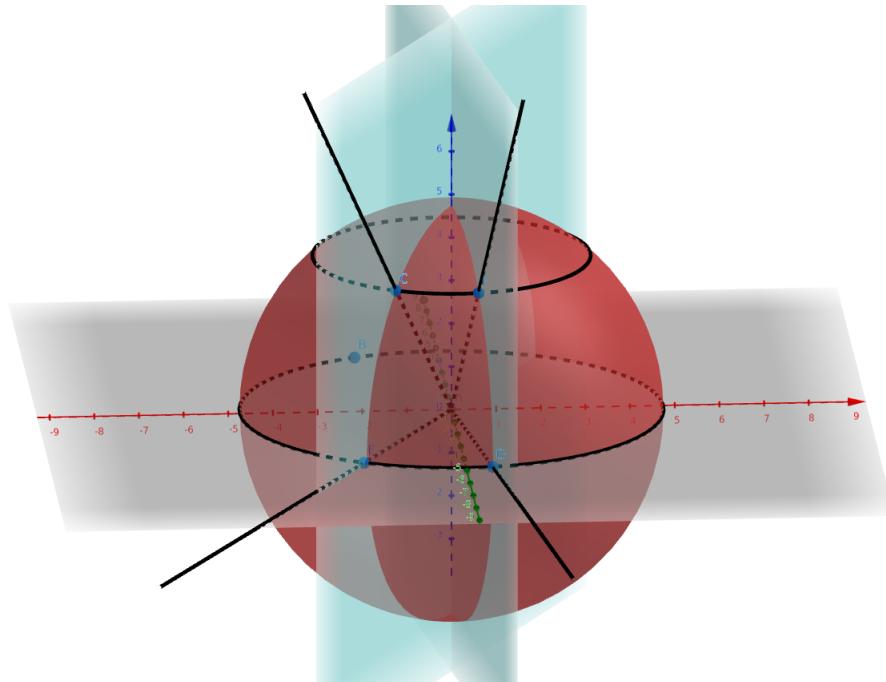


FIGURA 3. Medindo o ângulo de rotação.

Seja, então, u um vetor unitário ao longo do eixo e θ o ângulo da rotação ρ que queremos construir. Isto significa que

- $\rho(u) = u$;
- $\rho(w)$ e w formam um ângulo θ para todo vetor w perpendicular a u .

Para poder escrever a matriz de ρ , determinaremos uma base ortonormal $\{w_1, w_2\}$ do plano $W = \langle u \rangle^\perp$. Como o ângulo entre dois vetores não muda quando ambos são rodados por ρ , podemos concluir que $\rho(w_1)$ é perpendicular a u , de modo que, ao escrevê-lo como combinação linear da base

$$\beta = \{u, w_1, w_2\},$$

obtemos

$$\rho(w_1) = \langle \rho(w_1) | w_1 \rangle w_1 + \langle \rho(w_1) | w_2 \rangle w_2.$$

Para poder explicitar esta fórmula basta calcular

$$\langle \rho(w_1) | w_1 \rangle \text{ e } \langle \rho(w_1) | w_2 \rangle.$$

Como as rotações também não alteram o comprimento dos vetores, temos que $\rho(w_1)$ é unitário, de modo que os produtos internos acima são iguais aos ângulos entre os vetores. Mas o ângulo entre w_1 e $\rho(w_1)$ é θ pela definição da rotação, o que nos permite concluir que

$$\langle \rho(w_1) | w_1 \rangle = \cos(\theta).$$

Por outro lado, como w_2 é ortogonal a w_1 , o ângulo que forma com $\rho(w_1)$ é o que falta em θ para complementar $\pi/2$, donde

$$\langle \rho(w_1) | w_2 \rangle = \cos(\pi/2 - \theta) = \sin(\theta).$$

Assim, as coordenadas de $\rho(w_1)$ relativamente à base β são

$$(0, \cos(\theta), \sin(\theta))$$

Como w_2 é perpendicular a w_1 , um argumento semelhante mostra que

$$(0, -\cos(\pi/2 - \theta), \sin(\pi/2 - \theta)) = (0, -\sin(\theta), \cos(\theta)).$$

Portanto, a matriz de ρ relativamente à base β é

$$(\rho)_\beta = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{bmatrix}$$

Como, em geral, precisamos da matriz de ρ relativa à base canônica, teremos que aplicar mudança de base à matriz $(\rho)_\beta$ para obter $(\rho)_\varepsilon$. Para que isso seja possível, é necessário que conheçamos uma das duas matrizes que efetua a mudança de base entre β e ε . Como, na

maioria das aplicações, conhecemos as coordenadas dos vetores u , w_1 e w_2 relativamente à base canônica, vamos supor que é este o caso aqui, de modo que

$$(\text{id})_{\beta\varepsilon} = \begin{bmatrix} | & | & | \\ u & w_1 & w_2 \\ | & | & | \end{bmatrix}.$$

Note que os vetores de β têm que aparecer em $(\text{id})_{\beta\varepsilon}$ na mesma ordem em que foram listados em β . Finalmente, como vimos no artigo 5.5, o fato de β ser uma base ortonormal implica que a matriz $(\text{id})_{\beta\varepsilon}$ é ortogonal, o que nos permite escrever

$$(\rho)_\varepsilon = (\text{id})_{\beta\varepsilon}(\rho)_\beta(\text{id})_{\beta\varepsilon}^t.$$

Por exemplo, digamos que ρ é uma rotação cujo eixo é a reta de vetor diretor $(1, 1, 0)$ e o ângulo de rotação é $\pi/3$. Normalizando o vetor diretor do eixo, obtemos

$$u = \frac{1}{\sqrt{2}}(1, 1, 0)$$

de modo que podemos escolher

$$w_1 = \frac{1}{\sqrt{2}}(1, -1, 0) \text{ e } w_2 = (0, 0, 1),$$

como uma base ortonormal do plano perpendicular ao eixo. Portanto, segundo o que acabamos de ver, a matriz da rotação desejada relativamente à base

$$\beta = \left\{ \frac{1}{\sqrt{2}}(1, 1, 0), \frac{1}{\sqrt{2}}(1, -1, 0), (0, 0, 1) \right\}$$

será

$$(\rho)_\beta = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & -\sqrt{3}/2 \\ 0 & \sqrt{3}/2 & 1/2 \end{bmatrix}.$$

Para encerrar precisamos achar $(\rho)_\varepsilon$. Mas, pela definição da matriz de mudança de base,

$$(\text{id})_{\beta\varepsilon} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

de modo que

$$(\rho)_\varepsilon = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & -\sqrt{3}/2 \\ 0 & \sqrt{3}/2 & 1/2 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 3 & 1 & -\sqrt{6} \\ 1 & 3 & \sqrt{6} \\ \sqrt{6} & -\sqrt{6} & 2 \end{bmatrix}$$

3.4. Operadores autoadjuntos. Assim como no caso de dimensão dois, as projeções e reflexões do \mathbb{R}^n são exemplos de uma classe extremamente importante de operadores, os operadores autoadjuntos.

Um operador T do \mathbb{R}^n é *autoadjunto* se

$$\langle Tu | v \rangle = \langle u | Tv \rangle,$$

quaisquer que sejam os vetores $u, v \in \mathbb{R}^n$. Dito desta maneira, fica difícil identificar quais dos operadores que já conhecemos são autoadjuntos. O que precisamos é de uma tradução matricial da propriedade que define esta classe de operadores. Digamos que A é a matriz do operador T na base canônica. Então,

$$\langle Tu | v \rangle = (Au)^t v = u^t A^t v$$

ao passo que

$$\langle u | Tv \rangle = u^t (Av).$$

Como T é autoadjunto, teremos que

$$u^t A^t v = u^t (Av),$$

toda escolha de $u, v \in \mathbb{R}^n$, o que só pode acontecer se $A^t = A$. Portanto, como no caso do plano, os *operadores autoadjuntos são aqueles cuja matriz na base canônica é simétrica*. Uma afirmação análoga pode ser feita para a matriz do operador em qualquer base ortonormal. Para provar esta última afirmação, suponhamos que β seja uma outra base ortonormal do \mathbb{R}^n . Então,

$$(141) \quad (T)_\beta = Q^{-1}(T)_\varepsilon Q$$

em que escrevemos

$$Q = (\text{id})_{\beta\varepsilon}.$$

Mas as colunas de Q são os vetores de β , de modo que, pelo que vimos no artigo 5.5 do capítulo 6, a matriz Q é ortogonal. Isto nos permite reescrever (141) na forma

$$(T)_\beta = Q^t(T)_\varepsilon Q,$$

onde,

$$(T)_\beta^t = Q^t(T)_\varepsilon^t (Q^t)^t.$$

Como $(Q^t)^t = Q$ e $(T)_\varepsilon$ é simétrica por hipótese,

$$(T)_\beta^t = Q^t(T)_\varepsilon Q = (T)_\beta;$$

mostrando que $(T)_\beta$ também é uma matriz simétrica. Resumindo, temos o seguinte resultado.

PROPOSIÇÃO 3.1. *Um operador do \mathbb{R}^n é autoadjunto se, e somente se, sua matriz relativamente a alguma base ortonormal for simétrica.*

Como no caso do plano, é fácil usar esta proposição para mostrar que os operadores de projeção e reflexão do \mathbb{R}^n são autoadjuntos. Quando P é a matriz da projeção do \mathbb{R}^n sobre um subespaço U , temos que

$$P = u_1 u_1^t + \cdots + u_k u_k^t.$$

em que $\{u_1, \dots, u_k\}$ é uma base ortonormal de U . Segue-se, então, das propriedades da transposta que

$$P^t = (u_1^t)^t u_1^t + \cdots + (u_k^t)^t u_k^t = u_1 u_1^t + \cdots + u_k u_k^t = P.$$

No caso da reflexão, a demonstração é exatamente igual a que fizemos no artigo 2.2 do capítulo 1, por isso não vamos repeti-la aqui.

3.5. Construindo transformações lineares. Neste artigo veremos como obter alguns exemplos de transformações lineares com propriedades especificadas.

Suponhamos dados dois subespaços $U \subset \mathbb{R}^n$ e $W \subset \mathbb{R}^m$. Nossa primeira meta consiste em desenvolver um método para construir uma transformação linear $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ cujo núcleo é U e cuja imagem é W . Pelo teorema do núcleo e da imagem isto só é possível quando

$$(142) \quad n = \dim(U) + \dim(W).$$

Caso esta condição seja satisfeita, a construção executa os mesmos passos usados para provar o teorema do núcleo e da imagem. Assim, começamos achando uma base α para U , que completamos para uma base β do \mathbb{R}^n . A transformação desejada leva os vetores de α no zero e os vetores de $\beta \setminus \alpha$ nos vetores de uma base qualquer de W . Isto é possível porque (142) garante que

$$\#(\beta \setminus \alpha) = n - \dim(U) = \dim(W).$$

Por exemplo, digamos que queremos construir um operador linear T_1 do \mathbb{R}^4 cujo núcleo é o plano S gerado pelos vetores $(1, 1, 0, 0)$ e $(1, 0, 1, 1)$ e cuja imagem é o complemento ortogonal deste plano. Para obter o complemento ortogonal, completamos a base $\{(1, 1, 0, 0), (1, 0, 1, 1)\}$ de S para a base

$$\{(1, 1, 0, 0), (1, 0, 1, 1), (0, 0, 1, 0), (0, 0, 0, 1)\}$$

de \mathbb{R}^4 e aplicamos Gram-Schmidt, obtendo a base ortonormal formada pelos vetores

$$\begin{aligned} u_1 &= \frac{1}{\sqrt{2}}(1, 1, 0, 0) & u_2 &= \frac{1}{\sqrt{10}}(1, -1, 2, 2), \\ u_3 &= \frac{1}{\sqrt{15}}(-1, 1, 3, -2) & u_4 &= \frac{1}{\sqrt{3}}(-1, 1, 0, 1). \end{aligned}$$

Como

$$\langle u_1, u_2 \rangle = \langle (1, 1, 0, 0), (1, 0, 1, 1) \rangle$$

temos que $\{u_1, u_2\}$ é base ortonormal de S e S^\perp é gerado por $\langle u_3, u_4 \rangle$. Para definir T_1 , basta tomar

$$T_1(u_1) = T_1(u_2) = 0, \quad T_1(u_3) = u_3 \quad \text{e} \quad T_1(u_4) = u_4$$

que nos dá a matriz

$$(T_1)_\beta = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Como a matriz de mudança de base

$$(\text{id})_{\beta_1 \varepsilon} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{10}} & -\frac{1}{\sqrt{15}} & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{10}} & \frac{1}{\sqrt{15}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{10}} & \frac{3}{\sqrt{15}} & 0 \\ 0 & \frac{2}{\sqrt{10}} & -\frac{2}{\sqrt{15}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

é ortogonal, temos que

$$(T_1)_\varepsilon = Q(T_1)_\beta Q^t$$

onde

$$(T_1)_\varepsilon = \frac{1}{5} \begin{bmatrix} 2 & -2 & -1 & -1 \\ -2 & 2 & 1 & 1 \\ -1 & 1 & 3 & -2 \\ -1 & 1 & -2 & 3 \end{bmatrix}$$

Contudo, é importante observar que, ao contrário do que aconteceu neste exemplo, *raramente é possível construir uma base do domínio fazendo a união de uma base do núcleo com uma base da imagem*. Ilustraremos isto com segundo exemplo.

Suponhamos que queremos a matriz na base canônica de um operador linear T_2 do \mathbb{R}^3 cujo núcleo é o plano $U = \langle(1, 1, 1), (1, 0, -1)\rangle$ e cuja imagem é a reta $W = \langle(2, 1, 0)\rangle$. Como

$$(2, 1, 0) = (1, 1, 1) + (1, 0, -1),$$

o vetor $(2, 1, 0)$ é linearmente dependente dos geradores de U e, por isso, não pode ser usado para completá-lo para uma base do \mathbb{R}^3 . Isto não é um problema, porque a base do núcleo dada pode ser completada de qualquer maneira desejada. Por exemplo, podemos optar por acrescentar o vetor $(0, 0, 1) \notin U$ aos geradores de U , construindo, assim, a base

$$\alpha = \{(1, 1, 1), (1, 0, -1), (0, 0, 1)\}.$$

Para construir o operador desejado, tomamos

$$T_2(1, 1, 1) = 0, \quad T_2(1, 0, -1) = 0 \quad \text{e} \quad T_2(0, 0, 1) = (2, 1, 0).$$

Portanto, escolhendo α como a base do domínio e ε como a base do contra-domínio, obtemos a matriz

$$(T_2)_{\alpha\varepsilon} = \begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Como

$$(\text{id})_{\alpha\varepsilon} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & -1 & 1 \end{bmatrix} \quad \text{e} \quad (\text{id})_{\varepsilon\alpha} = (\text{id})_{\alpha\varepsilon}^{-1} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -1 & 0 \\ 1 & -2 & 1 \end{bmatrix}$$

temos, pela fórmula de mudança de base, que

$$(T_2)_\varepsilon = (T)_{\alpha\varepsilon}(\text{id})_{\varepsilon\alpha} = \begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & -1 & 0 \\ 1 & -2 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -4 & 2 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

que é a matriz desejada. Note que, embora não tenhamos feito isto, teria sido possível escolher β como uma base ortonormal, para isso bastaria aplicar o algoritmo de Gram-Schmidt aos três vetores que usamos para construir β .

Um outro caso em que não é possível completar os vetores do núcleo usando os vetores da imagem ocorre quando o domínio e o contradomínio da transformação são diferentes. Por exemplo, vejamos como determinar a matriz de uma transformação linear $T_3 : \mathbb{R}^4 \rightarrow \mathbb{R}^3$ cujo núcleo é o plano

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

e cuja imagem é o plano

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

Aplicando eliminação à matriz cujas linhas são os dois vetores que geram U , vemos que são linearmente independentes e que podemos completá-los acrescentando os vetores $(0, 1, 0, 0)$ e $(0, 0, 0, 1)$, obtendo, assim, a base

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

Fazendo

$$T(1, -1, 0, 0) = T(1, -1, 2, 3) = (0, 0, 0), \quad T(0, 1, 0, 0) = (1, 0, 1) \quad \text{e} \quad T(0, 0, 0, 1) = (2, 1, -1),$$

obtemos a matriz

$$(T_3)_{\gamma\varepsilon} = \begin{bmatrix} 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

Levando em conta que a matriz de mudança de base

$$(\text{id})_{\varepsilon\gamma} = (\text{id})_{\gamma\varepsilon}^{-1} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ -1 & -1 & 1 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 3 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & -\frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & -\frac{3}{2} & 1 \end{bmatrix}$$

temos que

$$(T_2)_\varepsilon = (T)_{\alpha\varepsilon} (\text{id})_{\varepsilon\alpha} = \begin{bmatrix} 1 & 1 & -3 & 2 \\ 0 & 0 & -\frac{3}{2} & 1 \\ 1 & 1 & \frac{3}{2} & -1 \end{bmatrix}$$

Outro problema que aparece frequentemente em questões práticas é o seguinte: *dados dois subespaços U e U' do \mathbb{R}^n , determinar um isomorfismo do \mathbb{R}^n nele próprio que leva U em U' .* Como isomorfismos preservam dimensão, este problema só tem solução quando $\dim(U) = \dim(U')$. Supondo que esta condição é verificada, escolhemos bases α para U e α' para U' que completamos para bases β e β' do \mathbb{R}^n , respectivamente. O isomorfismo desejado pode ser construído levando-se os vetores de β nos vetores de β' , de modo que aos vetores α sejam associados vetores α' .

Considere, por exemplo, as retas

$$U = \langle(1, 1, 1)\rangle \text{ e } U' = \langle(0, 1, 1)\rangle$$

do \mathbb{R}^3 . Como qualquer vetor na direção de $(1, 1, 1)$ gera U , preferimos trocá-lo por

$$\frac{1}{\sqrt{3}}(1, 1, 1)$$

na construção da base, porque isto nos permite acrescentar vetores de modo a obter uma base ortonormal de \mathbb{R}^3 ; por exemplo, podemos tomar

$$\beta = \left\{ \frac{1}{\sqrt{3}}(1, 1, 1), \frac{1}{\sqrt{2}}(0, 1, -1), \frac{1}{\sqrt{6}}(2, -1, -1) \right\}.$$

Procedendo de maneira semelhante para U' , escolhemos

$$\beta' = \left\{ \frac{1}{\sqrt{2}}(0, 1, 1), (1, 0, 0), \frac{1}{\sqrt{2}}(0, -1, 1) \right\}.$$

Para definir um isomorfismo $\psi : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ que leva U em U' por

$$\psi(1, 1, 1)/\sqrt{3} = (0, 1, 1)\sqrt{2}$$

$$\psi(0, 1, -1)/\sqrt{2} = (1, 0, 0)$$

$$\psi(2, -1, -1)/\sqrt{6} = (0, -1, 1)/\sqrt{2};$$

de modo que

$$(\psi)_{\beta\varepsilon} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & \sqrt{2} & 0 \\ 1 & 0 & -1 \\ 1 & 0 & 1 \end{bmatrix}$$

Note que pusemos nas colunas as coordenadas dos vetores de β' relativamente à base canônica, razão pela qual obtivemos $(\psi)_{\beta\varepsilon}$ e não $(\psi)_{\beta\beta'}$ como talvez você estivesse esperando. Para encontrar a matriz $(\psi)_{\varepsilon,\varepsilon}$ precisamos multiplicar $(\psi)_{\beta\varepsilon}$ à esquerda pela matriz de mudança de base $(\text{id})_{\varepsilon\beta}$. Para isto basta tomar a transposta de

$$(\text{id})_{\beta\varepsilon} = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & 0 & 2 \\ \sqrt{2} & \sqrt{3} & -1 \\ \sqrt{2} & -\sqrt{3} & 1 \end{bmatrix},$$

porque tivemos o cuidado de escolher β como sendo uma base ortonormal do \mathbb{R}^3 . Efetuando a multiplicação, encontramos

$$(\psi)_{\varepsilon,\varepsilon} = (\psi)_{\beta\varepsilon} (\text{id})_{\beta\varepsilon}^t = \frac{1}{12} \begin{bmatrix} 0 & 2\sqrt{3}\sqrt{6} & -2\sqrt{3}\sqrt{6} \\ (-2\sqrt{2}+2)\sqrt{6} & (\sqrt{2}+2)\sqrt{6} & (-\sqrt{2}+2)\sqrt{6} \\ (2\sqrt{2}+2)\sqrt{6} & (-\sqrt{2}+2)\sqrt{6} & (\sqrt{2}+2)\sqrt{6} \end{bmatrix}$$

Apesar de termos descrito a solução geral do problema em termos de uma reta, nada nos impede de considerar duas retas, em vez de uma, como origem e alvo da transformação. Adaptando um pouco o problema acima, podemos perguntar qual é o operador linear de \mathbb{R}^3 que leva as retas

$$U_1 = \langle(1, 1, 1)\rangle \text{ e } U_2 = \langle(0, 1, 0)\rangle,$$

respectivamente nas retas

$$U'_1 = \langle(0, 1, 1)\rangle \text{ e } U'_2 = \langle(1, 0, 0)\rangle.$$

A primeira coisa a observar é que os vetores que definem as retas U_1 e U_2 não são perpendiculares. Em particular, não podemos nos dar ao direito de escolher bases ortonormais ao resolver este problema. Mas a única razão para escolher estas bases é que simplificam as contas, nada mais. Como os vetores diretores das retas no domínio e contra-domínio não são colineares, os conjuntos

$$\{(1, 1, 1), (0, 1, 0)\} \text{ e } \{(0, 1, 1), (1, 0, 0)\}$$

são linearmente independentes e podemos completá-los a bases de \mathbb{R}^3 acrescentando um vetor a cada um. Como $(0, 0, 1)$ é independente de ambos os conjuntos, podemos acrescentá-lo a ambos, obtendo as bases

$$\beta = \{(1, 1, 1), (0, 1, 0), (0, 0, 1)\} \text{ e } \beta' = \{(0, 1, 1), (1, 0, 0), (0, 0, 1)\}.$$

O isomorfismo ψ fica então definido por

$$\psi(1, 1, 1) = (0, 1, 1)$$

$$\psi(0, 1, 0) = (1, 0, 0)$$

$$\psi(0, 0, 1) = (0, 0, 1)$$

onde obtemos a matriz

$$(\psi)_{\beta\varepsilon} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

Para obter a matriz $(\psi)_{\varepsilon\varepsilon}$ precisamos ainda inverter a matriz

$$(\text{id})_{\beta\varepsilon} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix},$$

já que, neste exemplo, ela não é ortogonal. Um cálculo fácil mostra que

$$(\text{id})_{\varepsilon\beta} = (\text{id})_{\beta\varepsilon}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix},$$

onde

$$(\psi)_{\varepsilon} = \begin{bmatrix} -1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

que é a matriz da transformação desejada.

Exercícios

1. Escreva a matriz correspondente a cada uma das transformações lineares dadas abaixo:
 - (a) $T : \mathbb{R}^4 \rightarrow \mathbb{R}^3$ definida por $T(x, y, z, w) = (x - y + z - w, x + y, 3z - 3w)$;
 - (b) $T : \mathbb{R}^3 \rightarrow \mathbb{R}^4$ definida por $T(x, y, z) = (x + y - z, x - y - 3z, x - 2z, y + z)$;
 - (c) $T : \mathbb{R}^4 \rightarrow \mathbb{R}^4$ definida por $T(x, y, z, w) = (x + y, x, -x + y, x + y)$;
 - (d) $T : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ definida por $T(x, y) = (x - y, x + y, x + y)$;
 - (e) $T : \mathbb{R}^4 \rightarrow \mathbb{R}^4$ definida por $T(x, y, z, w) = (x - y, z - w, x - y, z + w)$.
2. Escreva a matriz correspondente a cada uma das transformações lineares descritas abaixo:
 - (a) a projeção do \mathbb{R}^5 no hiperplano definido por $x - y - z + 2w + 5t = 0$;
 - (b) a reflexão do \mathbb{R}^3 cujo espelho é o plano de equação $x - 3y - 2z = 0$;

- (c) a rotação do \mathbb{R}^3 de eixo $(1, 1, 1)$ por um ângulo de $\pi/3$ radianos, no sentido definido pela regra da mão direita tomando-se v no sentido do polegar.
3. Determine uma base e a dimensão da imagem e do núcleo de cada uma das transformações lineares dadas abaixo:
- $T : \mathbb{R}^4 \rightarrow \mathbb{R}^3$ definida por $T(x, y, z, w) = (x - y + z - w, x + y, 3z - 3w)$;
 - $T : \mathbb{R}^3 \rightarrow \mathbb{R}^4$ definida por $T(x, y, z) = (x + y - z, x - y - 3z, x - 2z, y + z)$;
 - $T : \mathbb{R}^4 \rightarrow \mathbb{R}^4$ definida por $T(x, y, z, w) = (x + y, x, -x + y, x + y)$;
 - $T : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ definida por $T(x, y) = (x - y, x + y, x + y)$;
 - $T : \mathbb{R}^4 \rightarrow \mathbb{R}^4$ definida por $T(x, y, z, w) = (x - y, z - w, x - y, z + w)$.
4. Determine o núcleo e a imagem de uma rotação e de uma reflexão no plano.
5. Sejam B_1 e B_2 bases do \mathbb{R}^n e denote, respectivamente, por M_1 e M_2 as matrizes que mudam coordenadas nas bases B_1 e B_2 em coordenadas na base canônica. Mostre que: a matriz que muda coordenadas na base B_1 em coordenadas na base B_2 é igual a $M_2 M_1^{-1}$;
6. Sejam B_1 e B_2 bases, respectivamente, do \mathbb{R}^m e do \mathbb{R}^n . Mostre que se T é uma transformação linear de \mathbb{R}^m em \mathbb{R}^n , então a j -ésima coluna da matriz $(T)_{B_1 B_2}$ definida por
- $$(T)_{B_1 B_2}(v)_{B_1} = (Tv)_{B_2}$$
- é igual $(Tv_j)_{B_2}$, em que v_j é o j -ésimo vetor da base B_1 .
7. Sejam B_1 e B_2 bases, respectivamente, do \mathbb{R}^m e do \mathbb{R}^n e denote por M_1 e M_2 as matrizes que mudam coordenadas nas bases B_1 e B_2 em coordenadas na base canônica. Mostre que se T é uma transformação linear de \mathbb{R}^m em \mathbb{R}^n , então matriz $(T)_{B_1 B_2}$ definida no exercício anterior satisfaz
- $$M_2(T)_{B_1 B_2} M_1^{-1} = (T)_\varepsilon.$$
8. Considere as seguintes bases de \mathbb{R}^2 :
- $$B_1 = \{(1, 1), (-1, 1)\} \text{ e } B_2 = \{(2, 1), (1, -2)\},$$
- além da base canônica ε . Determine:
- as coordenadas de $(2, 3)$ em relação à base B_1 ;
 - as matrizes de mudança de base $(I)_{B_1 \varepsilon}$, $(I)_{\varepsilon B_1}$, $(I)_{B_2 \varepsilon}$ e $(I)_{\varepsilon B_2}$;
 - as coordenadas do vetor $(v)_{B_2} = (0, 2)$ relativamente às bases ε e B_1 .

9. Seja $B_1 = \{(1, 3), (2, 4)\}$. Determine B_2 , sabendo-se que é uma base de \mathbb{R}^2 e que a matriz de mudança de base

$$(I)_{B_1 B_2} = \begin{bmatrix} -7 & 6 \\ -11 & 8 \end{bmatrix}$$

10. Seja

$$B = \left\{ \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right), \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) \right\}$$

uma base de \mathbb{R}^2 e x' e y' os eixos correspondentes aos vetores de B .

- (a) Mostre que a matriz de mudança de base $(I)_{B\varepsilon}$ é ortogonal.
- (b) Determine o ângulo entre o primeiro vetor da base B e o vetor $e_1 = (1, 0)$.
- (c) Esboce a posição dos eixos x' e y' relativamente a x e y .
- (d) Determine as equações da reta $x + y = 2$ e da hipérbole $xy = 2$ na base B
- (e) Determine a equação da parábola $y' = (x')^2$ relativamente à base canônica e esboce seu gráfico.

11. Seja $B = \{(1, 0, 1), (0, 1, 1), (-1, 1, -1)\}$ uma base de \mathbb{R}^3 . Determine:

- (a) as matrizes de mudança de base $(\text{id})_{B\varepsilon}$ e $(\text{id})_{\varepsilon B}$;
- (b) a equação cartesiana, em relação às coordenadas na base B , do plano gerado pelos vetores $(1, 0, 1)$ e $(0, 1, 1)$.

12. Seja $\beta = \{v_1, \dots, v_n\}$ uma base do \mathbb{R}^n e considere o conjunto

$$\beta' = \{v_1 - v_2, v_1 - v_3, \dots, v_1 - v_n, v_1 + v_2 + \dots + v_n\}.$$

- (a) Mostre que β' também é uma base de \mathbb{R}^n .
- (b) Calcule a matriz de mudança de base de β para β' .

13. Determine a matriz de cada um dos operadores lineares de \mathbb{R}^2 relacionados abaixo na base canônica e na base $B = \{(1, 1), (1, 2)\}$:

- (a) $T(x, y) = (3x + 4y, 8x - y)$;
- (b) $T(v_1) = 3v_1 + 8v_2$ e $T(v_2) = 9v_1 - 6v_2$;
- (c) $T(v_1) = v_2$ e $T(v_2) = -v_1$.

14. Considere a seguinte base de \mathbb{R}^3 :

$$B = \{(1, 1, 1), (1, 1, 0), (1, 0, -1)\}.$$

Para cada uma das transformações lineares T dadas abaixo determine as matrizes $(T)_{B\varepsilon}$, $(T)_{\varepsilon B}$, $(T)_{\varepsilon\varepsilon}$ e $(T)_{BB}$:

- (a) $T(x, y, z) = (2x + 3y, z, 7x - z)$;
- (b) $T(x, y, z) = (4x + 2y + 6z, y + z, 3x - z)$;
- (c) $T(v_1) = v_1 + 4v_2 - v_3$, $T(v_2) = 2v_1 - v_3$ e $T(v_3) = 8v_2 + 5v_3$;
- (d) $T(v_1 - v_2 - v_3) = v_1 + v_2$, $T(v_2 + v_3) = v_3$ e $T(v_3) = 8v_2 + 5v_3$;

- (e) $T(v_1) = (1, 2, 0)$, $T(v_2) = (1, 0, 1)$ e $T(v_3) = (1, -1, 1)$;
 (f) $T(1, 1, -1) = v_1 + v_2$, $T(-1, 1, 1) = v_2 - v_3$ e $T(1, 0, 1) = v_1 - v_3$;
 (g) $T(v_1) = 2v_1$, $T(v_2) = 5v_2$ e $T(v_3) = -8v_3$.

15. Seja T uma transformação linear de \mathbb{R}^3 em \mathbb{R}^2 que satisfaz

$$T(1, 0, 0) = (2, 0), \quad T(0, 1, 0) = (1, 1) \text{ e } T(0, 0, 1) = (1, -1).$$

- (a) Determine $T(x, y, z)$.
 (b) Encontre um vetor $v \in \mathbb{R}^3$ tal que $T(v) = (3, 2)$.
 (c) O vetor v encontrado em (b) é único? Justifique sua resposta.

16. Determine uma transformação linear $T : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ que satisfaça

$$T(1, 2, -1) = (1, 0) \text{ e } T(2, 1, 4) = (0, 1).$$

Mostre que existem muitas transformações lineares que satisfazem estas restrições, mas que para todas elas a imagem de $(1, 1, 1)$ é a mesma. Justifique sua resposta.

17. Seja $T : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ uma transformação linear para a qual

$$T(1, 1, 2) = (1, 1), \quad T(0, 1, 0) = (0, -2) \text{ e } T(0, 1, 1) = (0, 0).$$

Determine:

- (a) $T(x, y, z)$;
 (b) uma transformação linear S de \mathbb{R}^2 em \mathbb{R}^3 tal que $S(1, 1) = (1, 1, 2)$ e $S(0, -2) = (0, 1, 0)$;
 (c) a transformação composta $S \circ T$.

18. O operador linear T de \mathbb{R}^3 definido por

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

é inversível? Em caso afirmativo, calcule a matriz na base canônica de T^{-1} e determine $T^{-1}(x, y, z)$.

19. Considere os operadores $D(\alpha, \beta)$ e $S(\gamma)$ de \mathbb{R}^2 cujas matrizes relativas à base canônica são

$$D(\alpha, \beta) = \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix} \text{ e } S(\gamma) = \begin{bmatrix} 1 & \gamma \\ 0 & 1 \end{bmatrix}$$

Prove que, dado um operador linear T de \mathbb{R}^2 , podemos determinar números reais α, β, γ e θ de modo que a matriz de T na base canônica pode ser escrita como um produto $\rho_\theta D(\alpha, \beta)S(\gamma)$, em que ρ_θ denota a rotação anti-horária de ângulo θ .

20. Dê exemplo de um operador linear T de \mathbb{R}^3 cuja imagem seja gerada por $(1, 0, 1)$ e $(1, 2, 2)$.

21. Dê exemplo de um operador linear T de \mathbb{R}^3 cujo núcleo seja a reta $\langle(1, -1, 1)\rangle$.
22. Dê exemplo de um operador linear T de \mathbb{R}^3 cujo núcleo seja gerado por $(1, -1, 1)$ e cuja imagem seja gerada por $(1, 0, 1)$ e $(1, 2, 2)$.
23. Dê exemplo de uma transformação linear T de \mathbb{R}^3 em \mathbb{R}^4 cuja imagem seja gerada por $(1, 2, 0, -4)$ e $(2, 0, -1, -3)$. Qual o núcleo do seu exemplo?
24. Dê exemplo uma transformação linear T de \mathbb{R}^3 em \mathbb{R}^4 cuja imagem seja gerada seja o plano de equações $x - y + z = y - w = 0$ e cujo núcleo seja a reta de equações $x - y + z = y - z = 0$.
25. Dê exemplo uma transformação linear T de \mathbb{R}^4 em \mathbb{R}^3 cuja imagem seja gerada seja o plano de equação $x - y + z = 0$ e cujo núcleo seja o plano gerador pelos vetores $(1, 0, 1, 1)$ e $(0, 1, -1, 1)$.
26. Para cada um dos itens abaixo, determine um operador linear *injetivo* de \mathbb{R}^3 que faz o que se pede:
 - (a) leva a reta $y - 3x = z = 0$ na reta $y - x = z = 0$;
 - (b) leva o plano $x + y - z = 0$ no plano $z = 0$;
 - (c) leva o plano $x - z = 0$ no plano $y = 0$;
 - (d) leva o plano $x + y + z = 0$ na reta $x - y = z = 0$;
 - (e) leva a reta $y - 3x = z = 0$ no plano $y - x = 0$.
27. Dê exemplo de uma transformação linear de \mathbb{R}^3 em \mathbb{R}^4 que aplica o plano $x + y - z = 0$ no plano $x - y = z - w = 0$.
28. Dê exemplo de um operador linear injetivo de \mathbb{R}^3 que aplica o plano $x - y - z = 0$ no plano $2x + y - z = 0$.
29. Seja T um operador de \mathbb{R}^2 cuja imagem está contida em seu núcleo. Mostre que existe uma base B de \mathbb{R}^2 relativamente à qual a matriz de T é da forma

$$\begin{bmatrix} 0 & \gamma \\ 0 & 0 \end{bmatrix} \text{ para algum número real } \gamma.$$
30. Seja S o subespaço do \mathbb{R}^3 gerado pelos vetores $(1, 0, 0)$, $(0, 1, 1)$ e $(1, -1, -1)$.
 - (a) Determine o complemento ortogonal S^\perp de S .
 - (b) Dê exemplo de uma transformação linear $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ que tem S como imagem e S^\perp como núcleo.
31. Seja T a reflexão do \mathbb{R}^3 através do plano $x + y = 0$.
 - (a) Encontre uma base ortonormal B relativa à qual T é diagonal.
 - (b) Calcule a matriz de T na base canônica.

32. Seja P a projeção ortogonal do \mathbb{R}^3 sobre o plano $3x + y - z = 0$.
- Determine o núcleo e a imagem de P .
 - Determine uma base ortonormal B relativa à qual a matriz de P é $\text{diag}(1, 1, 0)$.
 - Determine $(Pv)_B$ para os vetores cujas coordenadas relativamente à base canônica são $(1, 0, 0)$, $(0, 1, 0)$ e $(1, 1, 0)$.
 - Esboce o desenho, no plano $3x + y - z = 0$ da projeção do quadrado cujos vértices são $A = (0, 0, 0)$, $B = (1, 0, 0)$, $C = (0, 1, 0)$ e $D = (1, 1, 0)$.
33. Seja P a projeção ortogonal do \mathbb{R}^3 ao longo do vetor $v = (1, 1, 1)$.
- Determine o núcleo e a imagem de P .
 - Determine uma base ortonormal B relativa à qual a matriz de P é $\text{diag}(1, 1, 0)$.
 - Determine $(Pv)_B$ para os vetores cujas coordenadas relativamente à base canônica são $(1, 0, 0)$, $(0, 0, 1)$, $(0, 1, 0)$, $(0, 1, 1)$, $(1, 0, 1)$, $(1, 1, 0)$ e $(1, 1, 1)$.
 - Esboce o desenho, no plano de projeção de P , do cubo cujos vértices são os pontos acima e a origem.
34. Seja P a projeção do \mathbb{R}^4 ao longo do vetor $(1, 1, 1, 1)$. Determine as coordenadas de (x, y, z, w) numa base B relativamente à qual a matriz de P é $\text{diag}(1, 1, 1, 0)$.

CAPÍTULO 8

Diagonalização de operadores

Neste capítulo generalizaremos os conceitos de autovalores e autovetores, introduzidos para o caso do plano no capítulo 2, para operadores lineares do \mathbb{R}^n . Usaremos isso, então, para provar que operadores autoadjuntos são sempre diagonalizáveis e para descrever rotações no \mathbb{R}^3 e \mathbb{R}^3 .

1. Autovalores e autovetores

Além de introduzir autovalores e autovetores para operadores do \mathbb{R}^n e de calcular alguns exemplos, provaremos nesta seção um resultado segundo o qual um operador do \mathbb{R}^n que tem n autovalores distintos sempre admite uma base de autovetores.

1.1. Definições e primeiros exemplos. Seja T um operador linear do \mathbb{R}^n . Um vetor não nulo $v \in \mathbb{R}^n$ é um *autovetor* de T associado ao *autovalor* $\lambda \in \mathbb{R}$ se

$$T(v) = \lambda v.$$

Juntando o vetor nulo ao conjunto dos autovetores de T associados a λ obtemos o conjunto

$$V_\lambda = \{v \in \mathbb{R}^n \mid T(v) = \lambda v\}.$$

Como $T(v) = \lambda v$ equivale a $(T - \lambda \text{id})(v) = 0$, também podemos definir V_λ como o núcleo do operador $T - \lambda \text{id}$. Isto nos permite concluir que V_λ é um subespaço do \mathbb{R}^n , conhecido como o *autoespaço* do autovalor λ .

Quando um operador T do \mathbb{R}^n admite uma base formada por seus autovetores, dizemos que é *diagonalizável*. Para entender o porquê deste nome, digamos que v_1, \dots, v_n são autovetores linearmente independentes de T associados aos autovalores $\lambda_1, \dots, \lambda_n$. Assim,

$$\beta = \{v_1, \dots, v_n\}$$

é uma base de autovetores de T . Como

$$T(v_i) = \lambda_i v_i \text{ para todo } 1 \leq i \leq n,$$

temos que

$$(143) \quad (T)_\beta = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix},$$

é uma matriz diagonal. Observe que não estamos supondo que os λ 's sejam todos distintos; na verdade, nada impede que sejam todos iguais. Levando em conta que os operadores com os quais trabalhamos estão, quase sempre, definidos em coordenadas relativamente à base canônica, é conveniente reformular a definição do que significa um operador ser diagonalizável em termos puramente matriciais.

PROPOSIÇÃO 1.1. *Um operador T do \mathbb{R}^n é diagonalizável se, e somente se, existe uma matriz inversível M tal que*

$$M^{-1}(T)_\varepsilon M$$

é uma matriz diagonal.

DEMONSTRAÇÃO. Se T for diagonalizável, ele admite uma base de autovetores β . Segue-se, então, de (143) que

$$(\text{id})_{\beta\varepsilon}^{-1}(T)_\varepsilon(\text{id})_{\beta\varepsilon} = (T)_\beta$$

é uma matriz diagonal. Portanto, neste caso, a equação no enunciado da proposição é satisfeita por $M = (\text{id})_{\beta\varepsilon}$. Reciprocamente, se existir uma matriz inversível M tal que

$$M^{-1}(T)_\varepsilon M = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix},$$

então,

$$(T)_\varepsilon M = M \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

Logo, denotando por v_i a i -ésima coluna de M , temos que

$$(T)_\varepsilon v_i = \lambda_i v_i,$$

de modo que $\{v_1, \dots, v_n\}$ é uma base do \mathbb{R}^n formada por autovetores de T ; o que mostra que T é diagonalizável e conclui a demonstração. \square

Autovalores e autovetores aparecem frequentemente entre os operadores que já estudamos. Por exemplo, os vetores não nulos do núcleo de um operador linear que não é injetivo são autovetores associados ao autovalor zero. Um caso particular deste exemplo é dado pelas projeções ortogonais sobre um hiperplano H do \mathbb{R}^n . Neste caso, o autoespaço do autovalor zero é igual à reta normal a H ; ao passo que H é o autoespaço do autovalor um. Como as projeções, as reflexões sempre têm dois autovalores distintos, 1 e -1 . Para elas, o hiperplano H que faz o papel de espelho é o autoespaço de 1, ao passo que a reta que ortogonal a H é o autoespaço de -1 .

Nestes casos, a interpretação geométrica dos operadores nos ajudou a encontrar seus autovalores e autovetores. Mas, para tratar de operadores mais gerais, precisamos de um método sistemático. Suponhamos que a matriz de um operador T do \mathbb{R}^n relativamente a uma base α é conhecida. Se v for um autovetor de T associado ao autovalor λ , teremos que

$$A(v)_\alpha = \lambda(v)_\alpha,$$

em que $A = (T)_\alpha$. Denotando por I a matriz identidade $n \times n$, podemos reescrever esta equação na forma

$$(A - \lambda I)(v)_\alpha = 0.$$

Como a definição de autovetor requer que $v \neq 0$, a equação anterior implica que o sistema homogêneo cuja matriz é $A - \lambda I$ tem que ser indeterminado. Portanto, esta matriz terá, necessariamente, determinante zero quando λ for um dos autovalores de A . Contudo, $\det(A - \lambda I)$ é uma expressão polinomial em λ . Em outras palavras, se t for uma variável, então os autovalores de A serão raízes do polinômio

$$p_A(t) = \det(A - tI);$$

que, como já vimos no capítulo 2, é conhecido como *polinômio característico* de A .

Um problema com esta maneira de calcular os autovalores é que parece depender da escolha de base. Para ver que não é este o caso, digamos que α' seja uma outra base do \mathbb{R}^n . Denotando $(\text{id})_{\alpha' \alpha}$ por M , temos, pela fórmula de mudança de base, que

$$(144) \quad A = (T)_\alpha = M(T)_{\alpha'} M^{-1}.$$

Portanto,

$$p_A(t) = \det(A - tI) = \det(M \cdot (T)_{\alpha'} \cdot M^{-1} - tM \cdot I \cdot M^{-1}).$$

Pondo em evidência, M à esquerda e M^{-1} à direita, obtemos

$$p_A(t) = \det(M((T)_{\alpha'} - \lambda I)M^{-1}).$$

Levando em conta que o determinante de um produto de matrizes é igual ao produto dos determinantes destas matrizes, obtemos

$$p_A(t) = \det(M) \det((T)_{\alpha'} - \lambda I) \det(M^{-1}).$$

Contudo,

$$\det(M) \det(M^{-1}) = \det(MM^{-1}) = \det(I) = 1;$$

onde

$$p_A(t) = \det((T)_{\alpha'} - \lambda I);$$

o que mostra que o polinômio característico é independente da base escolhida para calculá-lo.

Uma vez que tenhamos obtido um autovalor λ de T , fica fácil calcular seu autoespaço que, como vimos acima, é o conjunto solução do sistema homogêneo $(A - \lambda I)X = 0$. Entretanto, ao contrário do que ocorre com os autovalores, o cálculo dos autovetores depende da base α na qual na qual a matriz A representa o operador T . Nem poderia ser de outra forma, porque um vetor só pode ser representado em forma de n -upla relativamente a uma dada base. Assim, por (144),

$$(M(T)_{\alpha'} M^{-1} - \lambda I)X = (A - \lambda I)X = 0;$$

que equivale a

$$M((T)_{\alpha'} - \lambda I)M^{-1}X = 0.$$

Como M é inversível,

$$((T)_{\alpha'} - \lambda I)M^{-1}X = 0.$$

Finalmente,

$$M^{-1}X = (\text{id})_{\alpha'} X$$

que representa X na base α' .

1.2. Mais exemplos. Nossa primeiro exemplo o operador T do \mathbb{R}^3 cuja matriz na base canônica é

$$A = (T)_e = \begin{bmatrix} 1 & 0 & 2 \\ -1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

O polinômio característico será

$$p_A(t) = \det A = \begin{bmatrix} 1-t & 0 & 2 \\ -1 & -t & 1 \\ 1 & 1 & 2-t \end{bmatrix} = -t^3 + 3t^2 + t - 3,$$

cujas raízes são $-1, 1$ e 3 . Portanto, T admite três autovalores distintos. Para descobrir os autovetores associados a -1 , devemos resolver o sistema homogêneo $(A - (-1)I)X = 0$ cuja matriz é

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

Aplicando eliminação gaussiana a esta matriz, e simplificando o resultado, obtemos

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

que corresponde ao sistema linear

$$\begin{aligned} x + z &= 0 \\ y + 2z &= 0 \end{aligned}$$

cujas soluções são dadas por

$$(x, y, z) = (-z, -2z, z) = z(-1, -2, 1).$$

Logo, os autovetores de A associados ao autovalor -1 são os múltiplos não nulos de $(-1, -2, 1)$. Portanto, o autoespaço associado a -1 é

$$V_{(-1)} = \langle(-1, -2, 1)\rangle.$$

Cálculos semelhantes mostram que

$$V_1 = \langle(1, -1, 0)\rangle \text{ e } V_3 = \langle(1, 0, 1)\rangle.$$

Logo,

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

é um conjunto formado por três autovetores de T . Contudo, para que T seja diagonalizável, é necessário que os três vetores de β sejam linearmente independentes. Em princípio, teríamos que verificar isto mas, como veremos na proposição 1.3 do próximo artigo, autovetores associados a autovalores distintos são necessariamente linearmente independentes. Portanto, β é uma base do \mathbb{R}^3 formada por autovetores de T , que nos permite afirmar que T é diagonalizável. Mais precisamente, temos que

$$(T)_\beta = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix} \quad \text{e} \quad (\text{id})_{\beta\varepsilon} = \begin{bmatrix} -1 & 1 & 1 \\ -2 & -1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

Note que, tanto a ordem em que os autovalores aparecem na matriz $(T)_\beta$, quando a ordem das colunas da matriz de mudança de base, têm que respeitar a ordem dos vetores na base β .

Seja, agora, o operador do \mathbb{R}^n que tem

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

como matriz na base canônica. Como J_n e, consequentemente, $J_n - tI$ são matrizes triangulares superiores, seus determinantes são iguais ao produto de suas entradas diagonais. Logo, o polinômio característico de J_n é

$$\det(J_n - tI) = \det \begin{bmatrix} 1-t & 1 & 0 & \cdots & 0 & 0 \\ 0 & 1-t & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1-t & 1 \\ 0 & 0 & 0 & \cdots & 0 & 1-t \end{bmatrix} = (1-t)^n,$$

de modo que 1 é o único autovalor de J_n . Para encontrar os autovetores associados a este autovalor precisamos achar as soluções do sistema

$$(J_n - I)X = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

que são dadas por

$$x_2 = x_3 = \cdots = x_{n-1} = x_n = 0.$$

Portanto, o autoespaço de 1 é a reta de vetor diretor $e_1 = (1, 0, \dots, 0, 0)$. Temos, assim, um operador que está longe de ser diagonalizável. A matriz J_n é conhecida como *matriz de Jordan* em homenagem ao matemático C. Jordan, que revelou o papel decisivo que desempenham na diagonalizabilidade dos operadores lineares.

Encerramos com um exemplo muito importante em várias aplicações, entre elas o algoritmo de busca originalmente utilizado pelo Google. Seja E uma matriz que não tem entradas negativas. Diremos que E é *estocástica por linhas* se a soma das entradas de cada uma de suas linhas é sempre igual a um. Uma matriz é *estocástica por colunas* se sua transposta é estocástica por linhas.

Se E for uma matriz estocástica por linhas de tamanho $n \times n$ e

$$u = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix}$$

então um cálculo fácil mosra que $Eu = u$. Assim, u é um autovetor de E associado ao autovalor um. Mas os autovalores de E são os mesmos de sua transposta, pois

$$\det(E - sI) = \det((E - sI)^t) = \det(E^t - sI).$$

Logo, uma matriz estocástica por colunas também tem 1 como autovalor, ainda que não possamos identificar facilmente seus autovetores. Resumindo, temos o seguinte resultado.

PROPOSIÇÃO 1.2. *Toda matriz estocástica (por linhas ou por colunas) tem 1 como um de seus autovalores.*

1.3. Duas proposições úteis. Ao afirmar que o operador do primeiro exemplo do artigo anterior era diagonalizável precisamos verificar (ainda que não tenhamos efetuado os cálculos) que os autovetores que determinamos eram linearmente independentes, do contrário não teríamos uma base. A proposição que provaremos neste artigo significa que nunca mais teremos que fazer isto.

PROPOSIÇÃO 1.3. *Autovetores associados a autovalores distintos têm que ser linearmente independentes.*

DEMONSTRAÇÃO. Seja T um operador linear do \mathbb{R}^n e suponhamos que $\lambda_1, \dots, \lambda_m$ são autovalores *distintos* de T e que v_1, \dots, v_m os autovetores associados a cada um destes autovalores. Provaremos por indução em j que o conjunto

$$L_j = \{v_1, \dots, v_j\}$$

é linearmente independente. Como autovetores têm que ser não nulos, $L_1 = \{v_1\}$ tem que ser linearmente independente, o que prova a base da indução. Suponha, agora, que L_j é linearmente independente. Usaremos isto para provar que L_{j+1} também é. Digamos que

$$(145) \quad a_1 v_1 + \cdots + a_{j+1} v_{j+1} = 0.$$

Calculando a imagem deste vetor por T temos que

$$a_1 T(v_1) + \cdots + a_{j+1} T(v_{j+1}) = 0.$$

Por outro lado, como v_i é autovetor associado a λ_i ,

$$0 = a_1 T(v_1) + \cdots + a_{j+1} T(v_{j+1}) = a_1 \lambda_1 v_1 + \cdots + a_{j+1} \lambda_{j+1} v_{j+1}.$$

Logo,

$$0 = \lambda_{j+1}(a_1v_1 + \cdots + a_{j+1}v_{j+1}) - (a_1\lambda_1v_1 + \cdots + a_{j+1}\lambda_{j+1}v_{j+1});$$

isto é,

$$a_1(\lambda_{j+1} - \lambda_1)v_1 + \cdots + a_j(\lambda_{j+1} - \lambda_j)v_j = 0.$$

Aplicando a hipótese de indução, podemos concluir que os coeficientes desta combinação linear têm que ser nulos; donde

$$a_1(\lambda_{j+1} - \lambda_1) = \cdots = a_j(\lambda_{j+1} - \lambda_j) = 0.$$

Como os λ 's são todos distintos,

$$\lambda_{j+1} - \lambda_i \neq 0 \text{ para todo } 1 \leq i \leq j.$$

Logo,

$$a_1 = \cdots = a_j = 0.$$

Substituindo isto em (145), resta apenas $a_{j+1}v_{j+1} = 0$. Como $v_{j+1} \neq 0$, temos também que $a_{j+1} = 0$. Logo, v_1, v_2, \dots, v_{j+1} são linearmente independentes. O resultado desejado segue pelo princípio de indução finita. \square

O seguinte corolário é a principal aplicação que faremos desta proposição.

COROLÁRIO 1.4. *Um operador do \mathbb{R}^n que tem n autovalores distintos tem que ser diagonalizável.*

Cada autovalor admite um autovetor associado, que a proposição garante serem linearmente independentes. Temos, assim, n autovetores linearmente independentes em \mathbb{R}^n , de modo que T tem que ser diagonalizável.

Quando o operador T é autoadjunto, a proposição acima e seu corolário dão lugar a um resultado ainda mais preciso.

PROPOSIÇÃO 1.5. *Autovetores associados a autovalores distintos de um operador autoadjunto têm que ser ortogonais.*

DEMONSTRAÇÃO. Seja T um operador autoadjunto do \mathbb{R}^n e suponhamos que v_1 e v_2 são autovetores de T associados aos autovalores λ_1 e λ_2 . Como T é autoadjunto,

$$(146) \quad \langle T(v_1) | v_2 \rangle = \langle v_1 | T(v_2) \rangle.$$

Contudo,

$$T(v_1) = \lambda_1 v_1 \quad \text{e} \quad T(v_2) = \lambda_2 v_2;$$

de modo que obtemos de (146) a igualdade

$$\lambda_1 \langle v_1 | v_2 \rangle = \lambda_2 \langle v_1 | v_2 \rangle.$$

Subtraindo um lado do outro,

$$(\lambda_1 - \lambda_2) \langle v_1 | v_2 \rangle = 0.$$

Portanto, se $\lambda_1 \neq \lambda_2$, então

$$\langle v_1 | v_2 \rangle = 0;$$

como queríamos provar. \square

Esta segunda proposição tem uma consequência análoga ao corolário 1.4 e igualmente fácil de provar.

COROLÁRIO 1.6. *Um operador autoadjunto do \mathbb{R}^n que tem n autovalores distintos admite uma base ortonormal.*

Este resultado é um caso especial do teorema espectral que, como no caso do plano, afirma que todo operador autoadjunto é diagonalizável, e que provaremos na próxima seção.

2. Teorema espectral

Nesta seção provamos o teorema espectral, segundo o qual todo operador autoadjunto é diagonalizável. Contudo, começaremos investigando uma generalização do conceito de autoespaço, que também será utilizada na caracterização das rotações na seção 3.

2.1. Subespaços invariantes. Os subespaços invariantes são uma generalização dos autoespaços introduzidos na seção anterior e serão utilizados, tanto na demonstração do teorema espectral, quanto na caracterização das rotações em \mathbb{R}^2 e \mathbb{R}^3 .

Sejam T um operador e U um subespaço do \mathbb{R}^n . Dizemos que U é *invariante* por T se todo vetor de U é levado por T em um vetor que também está em U . Isto é,

$$T(U) = \{T(u) \mid u \in U\} \subset U.$$

Como já observado acima, os autoespaços de um operador são exemplos de subespaços invariantes. De fato, se $\lambda \in \mathbb{R}$ for um dos autovalores de um operador T do \mathbb{R}^n e $v \in V_\lambda$, então

$$T(v) = \lambda v \in V_\lambda.$$

Outra ocasião em que nos defrontamos com o mesmo tipo de propriedade surgiu quando construímos rotações do \mathbb{R}^n no artigo 2.3 do capítulo 7. Nossa ponto de partida naquele artigo foi o teorema de Euler segundo o qual toda rotação ρ tem um eixo e os vetores do plano W ortogonal a seu eixo são levados em vetores que também pertencem a W . Esta última propriedade implica que, embora não seja um autoespaço de ρ , o plano W é invariante por ρ .

Antes de fazermos nossa primeira aplicação dos subespaços invariantes, precisamos de um resultado sobre matrizes. Dado um polinômio

$$f(t) = a_k t^k + \cdots + a_0$$

com coeficientes reais e uma matriz quadrada A , denotaremos por $f(A)$ a matriz obtida substituindo a variável t por A ; isto é,

$$f(A) = a_k A^k + \cdots + a_1 A + a_0 I,$$

em que I é a matriz identidade de mesmo tamanho que A . Por exemplo, Se

$$A = \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix} \quad \text{e} \quad f(t) = 2t^2 + 4t - 3,$$

então

$$f(A) = 2 \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix}^2 + 4 \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix} - 3 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -1 & -6 \\ 12 & -7 \end{bmatrix}.$$

Dizemos que $g_A(t)$ é o *polinômio mínimo* de A se for o polinômio de menor grau para o qual

$$g_A(A) = 0.$$

O próximo lema mostra que toda matriz tem um polinômio mínimo.

LEMA 2.1. *Dada uma matriz quadrada A , existe um polinômio não nulo $f(t)$ tal que $f(A) = 0$. Portanto, toda matriz admite um polinômio mínimo.*

DEMONSTRAÇÃO. Uma matriz $n \times n$ pode ser considerada como um vetor do \mathbb{R}^{n^2} , bastando, para isto, concatenar suas linhas uma em seguida à outra. Assim, à matriz

$$A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ a_{2,1} & \cdots & a_{2,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{bmatrix}$$

fazemos corresponder o vetor

$$\underbrace{(a_{1,1}, \dots, a_{1,n})}_{\text{linha 1}}, \underbrace{(a_{2,1}, \dots, a_{2,n}), \dots,}_{\text{linha 2}} \underbrace{(a_{n,1}, \dots, a_{n,n})}_{\text{linha } n}.$$

Isto nos permite aplicar às matrizes $n \times n$ tudo o que sabemos sobre vetores no \mathbb{R}^n . Em particular, pelo lema da troca, o conjunto

$$\{I, A, A^2, \dots, A^{n^2}\}$$

tem que ser linearmente dependente, porque contém $n^2 + 1$ matrizes. Logo, existem números reais $a_0, \dots, a_{n^2} \in \mathbb{R}$, nem todos nulos, tais que

$$a_{n^2} A^{n^2} + \cdots + a_0 I = 0.$$

Para obter o resultado do lema basta escolher k como sendo o maior índice entre 0 e n^2 para o qual $a_k \neq 0$ e tomar

$$f(t) = a_k t^k + \cdots + a_1 t + a_0.$$

O polinômio mínimo de A é o polinômio de menor grau entre todos os polinômios não nulos $f(t)$ para os quais $f(A) = 0$. \square

Antes de prosseguir, ilustraremos este resultado calculando o polinômio mínimo da matriz

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

Usando a correspondência da demonstração do lema, podemos considerar cada matriz 2×2 como um vetor do \mathbb{R}^4 . Como o conjunto formado por I , A e pelas matrizes

$$A^2 = \begin{bmatrix} -1 & -1 \\ 2 & -2 \end{bmatrix}, A^3 = \begin{bmatrix} -3 & 1 \\ -2 & -2 \end{bmatrix} \quad \text{e} \quad A^4 = \begin{bmatrix} -1 & 3 \\ -6 & 2 \end{bmatrix}$$

tem mais de quatro elementos, ele tem que ser linearmente dependente. Para calcular o polinômio mínimo, igualamos a zero a combinação linear

$$a_0I + a_1A + a_2A^2 + a_3A^3 + a_4A^4 = \begin{bmatrix} -a_4 - 3a_3 - a_2 + a_1 + a_0 & 3a_4 + a_3 - a_2 - a_1 \\ -6a_4 - 2a_3 + 2a_2 + 2a_1 & 2a_4 - 2a_3 - 2a_2 + a_0 \end{bmatrix}$$

obtendo um sistema linear cujas soluções são dadas por

$$a_0 = 2a_2 + 2a_3 - 2a_4 \quad \text{e} \quad a_1 = -a_2 + a_3 + 3a_4.$$

Assim, qualquer polinômio $f(t)$, de grau menor ou igual a 4, para o qual $f(A) = 0$ tem que ser da forma

$$f(t) = (2a_2 + 2a_3 - 2a_4) + (-a_2 + a_3 + 3a_4)t + a_2t^2 + a_3t^3 + a_4t^4.$$

Portanto, para obter o polinômio mínimo de A basta tomar $a_2 = 1$ e $a_3 = a_4 = 0$, o que nos dá

$$g_A(t) = t^2 - t + 2.$$

Com isso podemos introduzir um novo subespaço invariante associado a um operador T do \mathbb{R}^n .

PROPOSIÇÃO 2.2. *Se $p(t)$ é um fator não constante do polinômio mínimo de A , então o conjunto*

$$W_p = \{v \in \mathbb{R}^n \mid p(A)v = 0\}$$

é um subespaço não nulo do \mathbb{R}^n . Além disso, se $p(t) = t - \lambda$, para algum $\lambda \in \mathbb{R}$, então λ é autovalor de T e $W_p = V_\lambda$.

DEMONSTRAÇÃO. Como $p(A)$ é uma matriz, W_p é o conjunto solução do sistema $p(A)X = 0$ e, portanto, é um subespaço do \mathbb{R}^n . Para mostrar que este subespaço é invariante note, primeiramente, que se

$$p(t) = a_k t^k + \cdots + a_0,$$

então

$$p(A)A = a_k A^{k+1} + \cdots + a_0 A = Ap(A).$$

Logo, se $v \in W_p$ temos, pela associatividade do produto de matrizes, que

$$p(A) \cdot Av = p(A)A \cdot v = Ap(A) \cdot v = A \cdot p(A)v = 0,$$

em que a última igualdade é consequência da definição de W_p . Resta-nos provar que $W_p \neq \{0\}$. Como $p(t)$ é fator do polinômio mínimo $f(t)$ de A , existe um polinômio $h(t)$ tal que

$$f(t) = p(t)h(t).$$

Observe que $h(t)$ não pode ser nulo porque, por definição, o polinômio mínimo de uma matriz não é nulo. Mas, qualquer que seja o vetor $v \in \mathbb{R}^n$,

$$0 = f(A)v = p(A)h(A)v,$$

de modo que $h(A)v \in W_p$. Se $h(t)$ for uma constante o resultado é óbvio; caso contrário, $h(t)$ tem grau menor que o de $f(t)$ porque, por hipótese, $p(t)$ não é uma constante. Logo, pela minimalidade de $f(t)$, temos que $h(A) \neq 0$ que, por sua vez, implica que existe $v \in \mathbb{R}^n$ tal que $h(A)v \neq 0$, o que completa a demonstração da primeira parte da proposição. Para provar a segunda parte, observe que se $p(t) = t - \lambda$ então, pela primeira parte, existe $v \neq 0$ tal que

$$p(A)v = (A - \lambda I)v = 0;$$

donde

$$W_p = N(T - \lambda \text{id}) = V_\lambda.$$

□

A proposição acima tem como consequência o fato de que todo operador autoadjunto tem um autovalor real. A demonstração deste resultado, assim como a proposição anterior, são baseadas no artigo [6]. Na demonstração do lema usaremos um resultado, que provavelmente já é seu conhecido, segundo o qual *todo polinômio de grau maior ou igual a dois (com coeficientes reais) admite um fator de grau dois (com coeficientes reais)*.

LEMA 2.3. *Um operador autoadjunto sempre tem um autovalor real.*

DEMONSTRAÇÃO. Sejam T um operador autoadjunto do \mathbb{R}^n , A sua matriz na base canônica e $g_A(t)$ o polinômio mínimo de A . Seja $p(t)$ um fator (com coeficientes reais) de $g_A(t)$. Pela proposição 2.2, existe um vetor não nulo $v \in \mathbb{R}^n$ tal que $p(A)v = 0$. Como observado no parágrafo que antecede o enunciado deste lema, podemos supor que se $p(t)$ não for linear, então é um polinômio quadrático sem raiz real. Dividindo $p(t)$ pelo coeficiente de seu termo de maior grau, podemos escrever

$$p(t) = t + a \text{ ou } p(t) = t^2 + bt + c.$$

No primeiro caso, $-a \in \mathbb{R}$ é um autovalor de T pela segunda parte da proposição 2.2. Portanto, para concluir a demonstração, basta mostrar que o segundo caso não pode ocorrer. Quando $p(t)$ tem grau dois,

$$A^2v + bAv + cv = p(A)v = 0;$$

de modo que, pelas propriedades do produto interno.

$$\langle A^2v | v \rangle + b\langle Av | v \rangle + c\langle v | v \rangle = 0.$$

Mas, usando que T é autoadjunto, esta expressão pode ser reescrita como

$$\langle Av | Av \rangle + b\langle Av | v \rangle + c\langle v | v \rangle = 0;$$

onde

$$\|Av\|^2 + c\|v\|^2 = -b\langle Av | v \rangle.$$

Elevando ambos os membros ao quadrado,

$$(\|Av\|^2 + c\|v\|^2)^2 = b^2\langle Av | v \rangle^2.$$

Mas, a desigualdade de Cauchy-Schwartz, nos dá que

$$|\langle Av | v \rangle| \leq \|Av\|\|v\|;$$

veja (111) da página 164. Assim,

$$(\|Av\|^2 + c\|v\|^2)^2 \leq b^2\|Av\|^2\|v\|^2.$$

Porém, como estamos supondo que $p(T)$ não tem raiz real, seu discriminante $b^2 - 4c$ tem que ser negativo; donde

$$(147) \quad (\|Av\|^2 + c\|v\|^2)^2 \leq b^2\|Av\|^2\|v\|^2 < 4c\|Av\|^2\|v\|^2$$

Contudo,

$$(148) \quad (\|Av\|^2 - c\|v\|^2)^2 + 4c\|Av\|^2\|v\|^2 = (\|Av\|^2 + c\|v\|^2)^2.$$

Combinando (147) e (148), obtemos

$$(\|Av\|^2 - c\|v\|^2)^2 + 4c\|Av\|^2\|v\|^2 < 4c\|Av\|^2\|v\|^2;$$

que equivale a

$$(\|Av\|^2 - c\|v\|^2)^2 < -3c\|Av\|^2\|v\|^2;$$

contradizendo o fato de que o quadrado de um número real é sempre maior ou igual a zero. Portanto, $p(t)$ não pode ter grau dois, completando a demonstração da proposição. \square

2.2. O teorema espectral. Resta-nos utilizar o que fizemos no artigo anterior para provar o teorema espectral.

TEOREMA ESPECTRAL. *Todo operador autoadjunto admite uma base ortonormal de autovetores e, portanto, é diagonalizável e todos os seus autovalores são reais.*

DEMONSTRAÇÃO. A demonstração fornece um algoritmo recursivo que poderia, em princípio, ser usado para diagonalizar um operador autoadjunto. Para descrever este algoritmo partiremos da matriz A que representa um dado operador autoadjunto T do \mathbb{R}^n relativamente à base canônica. Como mostramos no artigo anterior, T admite um autovalor real, que chamaremos de λ , ao qual corresponde um autovetor u , que podemos supor ser unitário. Isto nos permite completar u para uma base ortonormal β do \mathbb{R}^n , de modo que

$$Q = (\text{id})_{\beta\varepsilon}$$

é uma matriz ortogonal. Como $Tu = \lambda u$, a fórmula de mudança de base nos dá,

$$(149) \quad QAQ^t = (T)_\beta = \begin{bmatrix} \lambda & \alpha_2 & \dots & \alpha_n \\ 0 & C & & \\ \vdots & & & \\ 0 & & & \end{bmatrix}$$

em que os α 's são números reais e C é uma matriz $(n-1) \times (n-1)$. Transpondo a equação acima e levando em conta que A é simétrica, obtemos

$$(150) \quad QAQ^t = \begin{bmatrix} \lambda & 0 & \dots & 0 \\ \alpha_2 & C^t & & \\ \vdots & & & \\ \alpha_n & & & \end{bmatrix}.$$

Igualando as matrizes em (149) e (150), concluímos os α 's são todos nulos e que $C^t = C$. Assim,

$$QAQ^t = \begin{bmatrix} \lambda & 0 & \dots & 0 \\ 0 & C & & \\ \vdots & & & \\ 0 & & & \end{bmatrix}$$

Como C é simétrica de tamanho $(n-1) \times (n-1)$, podemos continuar recursivamente até construir uma base ortonormal de autovetores para T . \square

Encerraremos a seção calculando uma base ortonormal para o operador autoadjunto T do \mathbb{R}^4 cuja matriz na base canônica é

$$\frac{1}{9} \begin{bmatrix} 14 & -4 & 0 & 2 \\ -4 & 14 & 0 & 2 \\ 0 & 0 & 9 & 0 \\ 2 & 2 & 0 & 17 \end{bmatrix}.$$

Como esta matriz é simétrica, sabemos que T é autoadjunto e, portanto, diagonalizável. O polinômio característico da matriz A é

$$p(t) = t^4 - 6t^3 + 13t^2 - 12t + 4,$$

cujas raízes são 1 e 2, ambas com multiplicidade dois. Portanto, temos dois autoespaços, que são obtidos resolvendo os sistemas $(A - I)X = 0$ e $(A - 2I)X = 0$. Fazendo isto, obtemos os autovetores

$$(-2, -2, 0, 1), (0, 0, 1, 0),$$

associados ao autovalor um, e

$$(1, 0, 0, 2), (-1, 1, 0, 0)$$

associados ao autovalor dois. Os autovetores associados 1 já são ortogonais, de modo que basta normalizá-los para obter a base ortonormal

$$\beta_1 = \left\{ \frac{1}{3}(-2, -2, 0, 1), (0, 0, 1, 0) \right\}$$

do autoespaço de 1. Como os autovetores que calculamos para 2 não são ortogonais, usaremos Gram-Schmidt para obter uma base ortonormal. O primeiro vetor desta base será

$$u_1 = \left(\frac{\sqrt{5}}{5}, 0, 0, \frac{2\sqrt{5}}{5} \right).$$

Para obter o segundo vetor, normalizamos

$$w_2 = (-1, 1, 0, 0) + \frac{1}{\sqrt{5}} \left(\frac{\sqrt{5}}{5}, 0, 0, \frac{2\sqrt{5}}{5} \right) = \left(-\frac{4}{5}, 1, 0, \frac{2}{5} \right),$$

obtendo

$$u_2 = \left(-\frac{4\sqrt{5}}{15}, \frac{\sqrt{5}}{3}, 0, \frac{2\sqrt{5}}{15} \right)$$

de modo que

$$\beta_2 = \left\{ \left(\frac{\sqrt{5}}{5}, 0, 0, \frac{2\sqrt{5}}{5} \right), \left(-\frac{4\sqrt{5}}{15}, \frac{\sqrt{5}}{3}, 0, \frac{2\sqrt{5}}{15} \right) \right\}$$

é uma base ortonormal do autoespaço do autovalor 2. Como, pela proposição 1.5, autovetores associados a autovalores distintos têm que ser ortogonais, o conjunto

$$\beta = \beta_1 \cup \beta_2$$

é uma base ortonormal de autovetores de T . A matriz de mudança de base e a matriz de T correspondentes a esta base são

$$(id)_{\beta\varepsilon} = \begin{bmatrix} -\frac{2}{3} & 0 & \frac{1}{\sqrt{5}} & -\frac{4\sqrt{5}}{15} \\ -\frac{2}{3} & 0 & 0 & \frac{\sqrt{5}}{3} \\ 0 & 1 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{2}{\sqrt{5}} & \frac{2\sqrt{5}}{15} \end{bmatrix} \quad \text{e} \quad (T)_{\beta} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

,

3. Rotações

Nesta seção caracterizamos as rotações nos espaços vetoriais de dimensões três e quatro e descrevemos a maneira como são construídas.

3.1. Operadores ortogonais. Seja ρ uma rotação de \mathbb{R}^3 . Como as rotações são operadores lineares que preservam a norma e o ângulo entre vetores, devemos ter que

$$\langle \rho(u), \rho(v) \rangle = \langle u, v \rangle;$$

quaisquer que sejam os vetores $u, v \in \mathbb{R}^3$. Denotando por Q a matriz de ρ na base canônica, a equação acima pode ser reescrita na forma

$$(Qu)^t(Qv) = u^tv;$$

onde

$$u^t Q^t Q v = u^t v,$$

que só pode valer para todo $u, v \in \mathbb{R}^3$ se $Q^t Q = I$. Portanto, *as rotações do \mathbb{R}^3 são operadores ortogonais*. Contudo, a recíproca desta afirmação é falsa porque a matriz de uma reflexão também é ortogonal. Isto nos leva a perguntar: *que propriedades nos permitem identificar as rotações dentre as transformações ortogonais?*

A resposta a esta pergunta está relacionada à noção de orientação. Dizemos que três vetores não nulos $u, v, w \in \mathbb{R}^3$ têm *orientação positiva* se w está do mesmo lado do plano $\langle u, v \rangle$ que o produto vetorial $u \times v$; do contrário estes vetores têm orientação negativa. Em outras palavras, vetores orientados positivamente têm a mesma posição relativa que os vetores e_1, e_2, e_3 da base canônica, que satisfazem $e_1 \times e_2 = e_3$. Como as rotações não alteram a posição relativa entre os vetores, podemos afirmar que *as rotações preservam a orientação de qualquer trinca de vetores à qual forem aplicadas*.

Podemos traduzir o fato de u, v, w terem orientação positiva em termos do produto interno. Afinal, se dois vetores estão do mesmo lado de um plano que não os contém, então o ângulo entre eles tem que ser agudo. Isto, por sua vez, implica que o produto interno entre estes vetores será necessariamente positivo. Como a recíproca desta afirmação também é verdadeira, concluímos que *três vetores $u, v, w \in \mathbb{R}^3$ estão positivamente orientados se, e somente se,*

$$\langle (u \times v) | w \rangle > 0$$

O número $\langle (u \times v), w \rangle$ é conhecido como *produto misto* dos vetores u, v, w e denotado por $[u, v, w]$. Combinando o que vimos até aqui, podemos afirmar que se ρ é uma rotação do \mathbb{R}^3 , então

$$[\rho(u), \rho(v), \rho(w)] = [u, v, w].$$

Antes de poder aplicar esta propriedade, precisamos de uma maneira de calcular o produto misto a partir das coordenadas dos vetores. Mas, se

$$u = (u_1, u_2, u_3) \text{ e } v = (v_1, v_2, v_3),$$

então, utilizando o determinante formal usual para calcular o produto vetorial, obtemos

$$u \times v = (u_2 v_3 - u_3 v_2, -u_1 v_3 + u_3 v_1, u_2 v_1 - u_1 v_2)$$

de modo que se $w = (w_1, w_2, w_3)$, então

$$[u, v, w] = w_1(u_2 v_3 - u_3 v_2) - w_2(u_1 v_3 - u_3 v_1) + w_3(u_2 v_1 - u_1 v_2);$$

que é igual ao determinante da matriz

$$A = \begin{bmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{bmatrix} = \begin{bmatrix} | & | & | \\ u & v & w \\ | & | & | \end{bmatrix}.$$

Portanto, se Q for a matriz de ρ , então

$$[\rho(u), \rho(v), \rho(w)] = \det \begin{bmatrix} | & | & | \\ Qu & Qv & Qw \\ | & | & | \end{bmatrix}.$$

Contudo, a matriz cujo determinante está sendo calculado na fórmula acima é igual a QA . Portanto,

$$[\rho(u), \rho(v), \rho(w)] = \det(QA) = \det(Q) \det(A).$$

Como

$$[u, v, w] = \det(A),$$

o operador ortogonal ρ cuja matriz é Q só pode preservar a orientação dos vetores u, v, w se $\det(Q) > 0$. Contudo, se Q é ortogonal, então

$$\det(Q^t Q) = \det(Q^t) \det(Q) = \det(Q)^2 = 1;$$

de modo que se Q é ortogonal e preserva orientação, então $\det(Q) = 1$. A proposição abaixo reune, de maneira sistemática, tudo o que fizemos até aqui nesta seção.

PROPOSIÇÃO 3.1. *Uma matriz 3×3 que corresponde a uma rotação do \mathbb{R}^3 tem que ser ortogonal de determinante um.*

Entretanto, isto ainda não responde à nossa pergunta original, porque não sabemos se toda matriz ortogonal de determinante um é uma rotação. Para poder responder a esta questão investigaremos estas transformações em mais detalhes no próximo artigo. Antes disto, porém, devemos observar que, ao contrário do que ocorre no \mathbb{R}^2 , uma matriz ortogonal 3×3 de determinante -1 não descreve necessariamente uma reflexão. Considere, por exemplo, o operador do \mathbb{R}^3 cuja matriz relativamente à base canônica é

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

Como $A^t = A$, temos que

$$AA^t = A^2 = I;$$

de modo que A é uma matriz ortogonal. Por outro lado,

$$\det(A) = (-1)^3 = -1.$$

Contudo esta matriz *não* pode representar uma reflexão, porque, qualquer que seja o vetor $v \in \mathbb{R}^3$, temos $Av = -v$, o que impede que haja um espelho, que seria o plano correspondente ao autoespaço de 1.

3.2. Autovalores de matrizes ortogonais. Seja Q uma matriz ortogonal que, a esta altura, podemos supor que é de tamanho $n \times n$, e seja

$$\det(Q - \lambda I)$$

seu polinômio característico. Como $QQ^t = I$, então

$$\det(Q - \lambda I) = \det(Q - \lambda QQ^t) = \det(Q(I - \lambda Q^t)).$$

Como o determinante de um produto de matrizes é igual ao produto dos determinantes destas matrizes,

$$\det(Q - \lambda I) = \det(Q) \det(I - \lambda Q^t).$$

Supondo que $\det(Q) = 1$, resta

$$\det(Q - \lambda I) = \det(I - \lambda Q^t).$$

Contudo, o determinante da transposta de uma matriz é igual ao determinante da própria matriz, donde

$$\det(Q - \lambda I) = \det((I - \lambda Q^t)^t) = \det(-\lambda Q + I).$$

Contudo, para qualquer matriz B de tamanho $n \times n$ e qualquer escalar c , temos que

$$\det(cB) = c^n \det(B);$$

de modo que

$$\det(Q - \lambda I) = (-\lambda)^n \det\left(Q - \frac{1}{\lambda}I\right).$$

Mas, quando n é ímpar, $(-\lambda)^n = -\lambda^n$, de modo que

$$(151) \quad \det(Q - \lambda I) = -\lambda^n \det\left(Q - \frac{1}{\lambda}I\right).$$

Fazendo $\lambda = 1$, esta última equação nos dá

$$\det(Q - I) = -\det(Q - I);$$

que só é possível se $\det(Q - I) = 0$; isto é, se 1 for raiz do polinômio característico de Q . Portanto, 1 tem que ser autovalor de Q , o que só pode acontecer se houver um vetor não nulo $u \in \mathbb{R}^3$ tal que $Qu = u$. Temos, então, uma reta inteira $\langle u \rangle$ cujos vetores não são modificados pela ação de Q . Esta reta é o *eixo* em torno do qual se manifesta a ação de Q . Usando esta nomenclatura, podemos resumir o que fizemos neste artigo de maneira bastante sucinta.

PROPOSIÇÃO 3.2. *Em um espaço de dimensão ímpar, todo operador ortogonal de determinante igual a um tem ao menos um eixo.*

Como 3 é ímpar, podemos deduzir desta proposição que toda rotação do \mathbb{R}^3 admite um eixo.

 É muito importante você observar que a proposição acima se aplica apenas a espaços vetoriais de dimensão ímpar. Quando a dimensão do espaço é par, a equação (151) não nos permite concluir nada quando tomamos $\lambda = 1$. Isto não é um mero defeito da nossa demonstração pois, como veremos no artigo 3.5, uma rotação do \mathbb{R}^4 pode não ter nenhum eixo.

3.3. Caracterizando as rotações em dimensão três. Suponhamos que T seja um operador de \mathbb{R}^3 cuja matriz Q , relativamente à base canônica, é ortogonal de determinante igual a um. Já sabemos, do artigo anterior, que T tem 1 como um de seus autovalores. Seja u um autovetor de T associado ao autovalor um. Se $v \in \mathbb{R}^3$, então

$$\langle u | Tv \rangle = \langle Tu | Tv \rangle = u^t Q^t Qv.$$

Como Q é ortogonal,

$$\langle u | Tv \rangle = u^t v = \langle u | v \rangle.$$

Logo, u e v são perpendiculares se, e somente se, u e Tv também são. Portanto, *qualquer vetor do plano S ortogonal a u é levado por T em outro vetor de S*. Em outras palavras, o subespaço S é invariante por T ; veja 2.1. Completando u para uma base ortonormal

$$\beta = \{u, w_1, w_2\}$$

do \mathbb{R}^3 , os vetores w_1 e w_2 estarão necessariamente em S . Como u é autovetor de Q associado ao autovalor 1, temos que

$$Tu = u.$$

Por outro lado, como S é invariante por T ,

$$Tw_1 = a_1 w_1 + a_2 w_2 \text{ e } Tw_2 = b_1 w_1 + b_2 w_2.$$

Portanto,

$$(T)_\beta = \begin{bmatrix} 1 & 0 & 0 \\ 0 & a_1 & b_1 \\ 0 & a_2 & b_2 \end{bmatrix}.$$

Denotando por A a matriz 2×2 igual a

$$\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix}$$

podemos escrever

$$(T)_\beta = \begin{bmatrix} 1 & 0 \\ 0 & A \end{bmatrix}.$$

Como T é ortogonal,

$$I = (T)_\beta ((T)_\beta)^t = \begin{bmatrix} 1 & 0 \\ 0 & A \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & A^t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & AA^t \end{bmatrix}$$

de modo que AA^t tem que ser igual à matriz identidade 2×2 . Por outro lado,

$$1 = \det(T)_\beta = \det(A)$$

de modo que A é uma matriz ortogonal 2×2 de determinante um. Mas estas matrizes são nossas velhas conhecidas, e já sabemos que podem ser escritas na forma

$$A = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

para algum ângulo θ . Portanto, a matriz da transformação T na base β será

$$(152) \quad (T)_\beta = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

Geometricamente isto significa que T tem o autoespaço de 1 como eixo e realiza uma rotação no plano perpendicular a este eixo. Em outras palavras, T é uma rotação. Isto nos permite enunciar a caracterização das rotações do espaço tridimensional que estávamos procurando.

TEOREMA 3.3. *Um operador de \mathbb{R}^3 é uma rotação se, e somente se, sua matriz na base canônica é ortogonal e tem determinante igual a um.*

Como é fácil de verificar, a mesma afirmação se aplica à matriz de T em qualquer base *ortonormal* de \mathbb{R}^3 . Entretanto o resultado é falso se a base *não* for ortonormal. Contudo, este teorema não esgota tudo o que aprendemos sobre rotações na argumentação acima. De fato, vimos também que toda rotação de \mathbb{R}^3 :

- tem um eixo;
- efetua uma rotação no plano ortogonal ao eixo.

Estas últimas afirmações são muito úteis na construção de exemplos, como aliás já sabemos desde o artigo 3.3.

Encerraremos nosso estudo das rotações com mais um exemplo. Sendo $u = (1, 1, 1)$ como antes, construiremos a matriz da reflexão R do \mathbb{R}^3 cujo espelho é o plano ortogonal a u , que é dada por

$$R = I - 2uu^t = \frac{1}{3} \begin{bmatrix} 1 & -2 & -2 \\ -2 & 1 & -2 \\ -2 & -2 & 1 \end{bmatrix}.$$

Como esta é uma matriz ortogonal que não descreve uma rotação, $\det(R) = -1$. Mas, mudar o sinal de uma das linhas da matriz inverte o sinal do determinante, de modo que

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

tem que ser uma matriz ortogonal de determinante igual a um. Logo Q descreve uma rotação: quais são o eixo e o ângulo desta rotação? Para determinar o eixo, calculamos o autovetor do autovalor 1 resolvendo o sistema

$$\frac{1}{3} \begin{bmatrix} -4 & 2 & 2 \\ -2 & -2 & -2 \\ -2 & -2 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

cujas soluções satisfazem

$$x = y + z = 0$$

de modo que $u = (0, 1, -1)$ é um vetor diretor do eixo de Q . Para determinar o ângulo de rotação, escolheremos um vetor w do plano ortogonal a u , e calcularemos o ângulo entre w e Qw . Tomando $w = (1, 0, 0)$, temos que

$$Qw = \frac{1}{3}(-1, -2, -2);$$

onde

$$\langle w, Qw \rangle = w^t Qw = -\frac{1}{3}.$$

Logo, o ângulo de rotação θ satisfaz

$$\cos(\theta) = -\frac{1}{3}$$

pois w é um vetor unitário. O ângulo θ correspondente é aproximadamente igual a 1,9106 radianos.

3.4. Calculando eixo e ângulo eficientemente. No artigo anterior usamos autovetores para encontrar o eixo de uma rotação no \mathbb{R}^3 e um vetor no plano ortogonal ao eixo para determinar o ângulo. Contudo há métodos algébricos mais eficientes para calcular ambos.

Seja ρ a rotação cujo eixo e ângulo queremos determinar e Q sua matriz na base canônica. Se u é um vetor diretor do eixo de ρ , então

$$Qu = u \quad \text{e} \quad Q^t u = Q^{-1} u = u.$$

A matriz

$$A = \frac{1}{2}(Q - Q^t)$$

satisfaz

$$Au = \frac{1}{2}(Qu - Q^t u) = 0.$$

Portanto, u pertence ao núcleo do operador cuja matriz na base canônica é A . Mas

$$A^t = \left(\frac{1}{2}(Q - Q^t) \right)^t = \frac{1}{2}(Q^t - Q) = -A,$$

de modo que, se

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix}$$

então

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

que só é possível se

$$a_{i,i} = 0 \quad \text{e} \quad a_{i,j} = -a_{j,i}$$

para todo $1 \leq i < j \leq 3$. Fazendo as substituições necessárias, verificamos que

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

Como

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

o vetor

$$(153) \quad u_0 = (a_{3,2}, -a_{3,1}, a_{2,1})$$

só não estará sobre o eixo de ρ se o núcleo de A tiver dimensão maior que um ou $A = 0$. Para descobrir sob que condições o núcleo de A tem dimensão dois ou mais, precisamos analisar o sistema $AX = 0$. Aplicando eliminação à matriz A quando $a_{2,1} \neq 0$, verificamos que sua forma escada pode ser escrita na forma

$$\begin{bmatrix} a_{2,1} & 0 & -a_{3,2} \\ 0 & -a_{2,1} & -a_{3,1} \\ 0 & 0 & 0 \end{bmatrix}.$$

Como os vetores correspondentes às duas primeiras colunas desta matriz são linearmente independentes, a imagem de A tem dimensão dois neste caso. Por outro lado, quando $a_{2,1} = 0$, a matriz A é igual a

$$\begin{bmatrix} 0 & 0 & -a_{3,1} \\ 0 & 0 & -a_{3,2} \\ a_{3,1} & a_{3,2} & 0 \end{bmatrix}.$$

Mas a imagem desta matriz tem dimensão dois a não ser que $a_{3,1} = a_{3,2} = 0$. Logo, a imagem de A tem dimensão dois sempre que $A \neq 0$. Portanto, pelo teorema do núcleo e da imagem, o núcleo de A tem dimensão um quando $A \neq 0$, garantindo que, neste caso, u_0 é sempre vetor diretor do eixo.

Por outro lado, quando $A = 0$ temos que $Q = Q^t$; de modo que, neste caso, Q é ortogonal e simétrica. Como estas duas propriedades continuam valendo se trocarmos da base canônica para outra base ortonormal qualquer, podemos supor β é a base formada pelo vetor diretor u do eixo e pelos dois vetores de uma base ortonormal de $\langle u \rangle^t$. Mas, para que

$$(\rho)_\beta = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{bmatrix}$$

seja simétrica é necessário que

$$\operatorname{sen}(\theta) = -\operatorname{sen}(\theta);$$

isto é, que $\operatorname{sen}(\theta) = 0$, o que só é possível se θ for múltiplo inteiro de π . Assim, as rotações ρ para as quais $A = 0$ têm matriz na base β igual a I ou a $-I$. Logo, em ambos os casos

$$(\rho)_\varepsilon = \pm (\operatorname{id})_{\beta\varepsilon} \cdot I \cdot (\operatorname{id})_{\varepsilon\beta} = \pm I.$$

Realisticamente falando, não estaríamos buscando o ângulo de rotação em nenhum destes dois casos. Portanto, para todos os propósitos práticos, podemos usar a receita encontrada acima para encontrar o eixo de qualquer rotação do plano.

O ângulo de rotação também pode ser calculado de maneira bastante simples usando o traço da matriz Q . Pela equação (57) da página 82, temos que

$$\operatorname{tr}(Q) = \operatorname{tr}((\operatorname{id})_{\beta\varepsilon}(\rho)_\beta(\operatorname{id})_{\beta\varepsilon}^{-1}) = \operatorname{tr}((\rho)_\beta).$$

Supondo que a base ortonormal β foi escolhida de modo que

$$(\rho)_\beta = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\operatorname{sen}(\theta) \\ 0 & \operatorname{sen}(\theta) & \cos(\theta) \end{bmatrix}$$

podemos concluir que

$$\operatorname{tr}(Q) = 1 + 2\cos(\theta).$$

Portanto,

$$(154) \quad \cos(\theta) = \frac{\operatorname{tr}(Q) - 1}{2}.$$

Encerramos aplicando o que aprendemos para achar o eixo da rotação, considerada ao final do artigo anterior, cuja matriz na base canônica é

$$Q = \frac{1}{3} \begin{bmatrix} -1 & 2 & 2 \\ -2 & 1 & -2 \\ -2 & -2 & 1 \end{bmatrix}.$$

Uma simples subtração nos dá

$$A = \frac{1}{2}(Q - Q^t) = \begin{bmatrix} 0 & \frac{2}{3} & \frac{2}{3} \\ -\frac{2}{3} & 0 & 0 \\ -\frac{2}{3} & 0 & 0 \end{bmatrix}$$

de modo que, por (153),

$$u_0 = \frac{1}{3}(0, 2, -2)$$

é um vetor diretor do eixo de Q . Finalmente, aplicando a fórmula (154), obtemos

$$\cos(\theta) = \frac{\frac{1}{3} - 1}{2} = -\frac{1}{3}$$

que é o mesmo resultado encontrado, com um esforço bastante maior, ao final do artigo 3.3. Uma demonstração do teorema de Euler sobre a existência do eixo de rotações no \mathbb{R}^3 que usa as ideias exploradas neste artigo pode ser encontrada em [5].

3.5. Caracterizando as rotações em dimensão quatro. Nos artigos anteriores utilizamos nosso conhecimento intuitivo de qual deveria ser o comportamento geométrico de uma rotação para caracterizar, no teorema 3.3, as matrizes das rotações no \mathbb{R}^3 . Naturalmente não podemos fazer nada semelhante em dimensão maior, porque não conseguimos visualizar estes espaços. Por isso, procedemos por analogia. Já que, tanto em dimensão dois, quanto em dimensão três, uma rotação é caracterizada como um operador cuja matriz é ortogonal e de dimensão zero, *definimos* uma rotação em dimensão n como sendo um operador do \mathbb{R}^n cuja matriz na base canônica é ortogonal e tem determinante um. Neste artigo usaremos esta definição para dar uma caracterização, semelhante à obtida no artigo anterior, para as matrizes das rotações no \mathbb{R}^4 .

Como vimos no artigo 3.2, toda rotação do \mathbb{R}^3 admite um eixo. Embora isto não se aplique às rotações do \mathbb{R}^4 , todas elas admitem dois planos invariantes. Porém, antes de provar isto, precisamos de um lema.

LEMA 3.4. *Se T é um operador ortogonal do \mathbb{R}^n e W é um subespaço invariante por T , então W^\perp também é invariante por T .*

DEMONSTRAÇÃO. Como W é invariante por T , temos que $T(W) \subset W$. Mas T é inversível, de modo que $T(W)$ tem que ter a mesma dimensão que W . Assim, pelo teorema 4.3, $T(W) = W$. Por outro lado, como T é ortogonal, segue-se do artigo 3.1, que

$$(155) \quad \langle Tv | Tw \rangle = \langle v | w \rangle.$$

Logo, se $w \in W$ e $v \in W^\perp$, então

$$\langle Tv | Tw \rangle = 0.$$

Contudo, como $T(W) = W$, todo elemento de W pode ser escrito na forma $T(w)$ para alguma $w \in W$, o que nos permite concluir que Tv é ortogonal a todos os elementos de W . Portanto, W^\perp é invariante por T , como queríamos mostrar. \square

PROPOSIÇÃO 3.5. *Todo operador ortogonal do \mathbb{R}^4 admite um plano invariante U . Além disso, o complemento ortogonal de U também é invariante pelo mesmo operador.*

DEMONSTRAÇÃO. Seja T um operador ortogonal do \mathbb{R}^n e Q sua matriz da base canônica. A demonstração consiste de dois casos, dependendo se T tem ou não algum autovalor real.

Suponhamos, primeiramente, que T não tem nenhum autovalor real e seja $g_Q(t)$ o polinômio mínimo de Q , como definido no artigo 2.1. Pela proposição 2.2, as raízes de $g_Q(t)$ correspondem aos autovalores de T . Portanto, $g_Q(t)$ tem que ter pelo menos um fator quadrático

$$p(t) = t^2 + bt + c,$$

que não admite nenhuma raiz real. Neste caso,

$$U = \langle v, Tv \rangle = \langle v, Qv \rangle,$$

é invariante por T , pois

$$(Q^2 + bQ + cI)v = 0$$

implica que

$$Q(Qv) = Q^2v = -bQv - cv \in U.$$

Além disso, como T não tem autovalor real, sabemos que v e que $T(v)$ não podem ser colineares, de modo que $\dim(U) = 2$.

Vejamos, agora, o que acontece quando T admite um autovalor real λ_1 , ao qual está associado um autovetor v_1 . Vimos no artigo 6.4 do capítulo 6 que se $S = \langle v_1 \rangle$, então

$$\mathbb{R}^4 = S \oplus S^\perp.$$

Mas, pelo lema 3.4, S^\perp é invariante por T . Como $\dim(S^\perp) = 3$, o polinômio característico do operador $T|_S$, obtido restringindo T a S^\perp , tem grau três. Contudo, todo polinômio de grau três com coeficientes reais admite uma raiz real. Logo, $T|_S$ admite um autovalor real λ_2 ao qual está associado um autovetor v_2 . Por definição da restrição,

$$T(v_2) = T|_S(v_2) = \lambda_2 v_2,$$

de modo que v_2 também é autovetor de T . Além disso, como v_2 é ortogonal a v_1 , eles geram um plano U . Se

$$a_1 v_1 + a_2 v_2 \in U,$$

para alguma escolha de $a_1, a_2 \in \mathbb{R}$, então

$$T(a_1 v_1 + a_2 v_2) = a_1 T(v_1) + a_2 T(v_2) = a_1 \lambda_1 v_1 + a_2 \lambda_2 v_2 \in U,$$

o que mostra que U é invariante por T . A segunda parte da demonstração é consequência imediata do lema 3.4. \square

Com isso podemos proceder à desejada caracterização das rotações do \mathbb{R}^4 . Seja, então, ρ uma rotação no \mathbb{R}^4 e $Q = (\rho)_\varepsilon$ sua matriz na base canônica. Lembre-se que, por definição, ρ é uma rotação se, e somente se, Q é ortogonal e tem determinante igual a um. A proposição 3.5 nos permite afirmar que existe um subespaço U , de dimensão dois, no \mathbb{R}^4 que é invariante por Q e cujo complemento ortogonal U^\perp também é invariante por Q . Sejam,

$$\beta_1 = \{u_1, u_2\} \quad \text{e} \quad \beta_2 = \{w_1, w_2\}$$

bases ortonormais de U e U^\perp , respectivamente. Como todo vetor de U é ortogonal a todo vetor de U^\perp ,

$$\beta = \beta_1 \cup \beta_2 = \{u_1, u_2, w_1, w_2\}$$

é uma base ortonormal do \mathbb{R}^4 . Mas, U e U^\perp são invariantes por ρ , de modo que

$$(156) \quad (\rho)_\beta = \begin{bmatrix} Q_1 & 0 \\ 0 & Q_2 \end{bmatrix}$$

em que cada bloco é uma matriz 2×2 . Contudo,

$$(157) \quad (\rho)_\beta = (\text{id})_{\beta\varepsilon}^t Q (\text{id})_{\varepsilon\beta},$$

de modo que $(\rho)_\beta$ também é uma matriz ortogonal. Por outro lado, por (157)

$$\det((\rho)_\beta) = \det(Q)$$

e por (156)

$$\det((\rho)_\beta) = \det(Q_1) \det(Q_2);$$

onde

$$\det(Q_1) \det(Q_2) = \det(Q).$$

Logo, $\det(Q) = 1$ se, e somente se, Q_1 e Q_2 são ambas rotações planas ou ambas reflexões planas. No primeiro caso, temos, pelo artigo 2.3, que existem ângulos θ_1 e θ_2 tais que

$$Q_1 = \begin{bmatrix} \cos(\theta_1) & -\sin(\theta_1) \\ \sin(\theta_1) & \cos(\theta_1) \end{bmatrix} \quad \text{e} \quad Q_2 = \begin{bmatrix} \cos(\theta_2) & -\sin(\theta_2) \\ \sin(\theta_2) & \cos(\theta_2) \end{bmatrix};$$

onde,

$$(158) \quad (\rho)_\beta = \begin{bmatrix} \cos(\theta_1) & -\sin(\theta_1) & 0 & 0 \\ \sin(\theta_1) & \cos(\theta_1) & 0 & 0 \\ 0 & 0 & \cos(\theta_2) & -\sin(\theta_2) \\ 0 & 0 & \sin(\theta_2) & \cos(\theta_2) \end{bmatrix},$$

que desempenha o mesmo papel para rotações do \mathbb{R}^4 que (152) desempenha para rotações do \mathbb{R}^3 . Por outro lado, se Q_1 e Q_2 são reflexões, podemos escolher β_1 e β_2 de modo que

$$Q_1 = Q_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Mas isto significa que

$$\rho(u_1) = u_1, \quad \rho(u_2) = -u_2, \quad \rho(w_1) = w_1, \quad \text{e} \quad \rho(w_2) = -w_2.$$

Seja

$$\beta' = \{u_1, w_1, u_2, w_2\}$$

a base obtida trocando de posição os vetores u_2 e w_1 da base β . Como esta base é ortonormal, temos que

$$\langle u_1, w_1 \rangle^\perp = \langle u_2, w_2 \rangle.$$

Por sua vez, a matriz de ρ nesta base é

$$(\rho)_{\beta'} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix},$$

que corresponde a (158) quando $\theta_1 = 0$ e $\theta_2 = -1$. Reunindo tudo o que fizemos, temos o seguinte resultado.

TEOREMA 3.6. *Se ρ é um operador do \mathbb{R}^4 cuja matriz na base canônica é ortogonal e tem determinante igual a 1, então existe uma base ortonormal β do \mathbb{R}^4 e ângulos θ_1 e θ_2 , tais que*

$$(\rho)_\beta = \begin{bmatrix} \cos(\theta_1) & -\sin(\theta_1) & 0 & 0 \\ \sin(\theta_1) & \cos(\theta_1) & 0 & 0 \\ 0 & 0 & \cos(\theta_2) & -\sin(\theta_2) \\ 0 & 0 & \sin(\theta_2) & \cos(\theta_2) \end{bmatrix}.$$

Reciprocamente, qualquer matriz desta forma é ortogonal e tem determinante 1.

Por exemplo, para construir a rotação ρ que gira de um ângulo $\pi/4$ no plano U gerado pelos vetores

$$u_1 = \frac{1}{2}(1, -1, -1, 1) \quad \text{e} \quad u_2 = \frac{1}{2}(-1, 1, -1, 1)$$

e de um ângulo igual a $\pi/3$ no plano U^\perp , gerado por

$$w_1 = \frac{1}{2}(-1, -1, 1, 1) \quad \text{e} \quad w_2 = \frac{1}{2}(1, 1, 1, 1)$$

começamos montando a base ortonormal

$$\beta = \{u_1, u_2, w_1, w_2\}$$

na qual ρ tem por matriz

$$(\rho)_\beta = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ 0 & 0 & \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}$$

Como

$$(\text{id})_{\beta\varepsilon} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

temos que

$$(\rho)_\beta = (\text{id})_{\beta\varepsilon}(\rho)_\beta(\text{id})_{\beta\varepsilon}^t = \begin{bmatrix} \frac{\sqrt{2}+1}{4} & \frac{-\sqrt{2}+1}{4} & \frac{\sqrt{3}+\sqrt{2}}{4} & \frac{\sqrt{3}-\sqrt{2}}{4} \\ \frac{-\sqrt{2}+1}{4} & \frac{\sqrt{2}+1}{4} & \frac{\sqrt{3}-\sqrt{2}}{4} & \frac{\sqrt{3}+\sqrt{2}}{4} \\ \frac{-\sqrt{3}-\sqrt{2}}{4} & \frac{-\sqrt{3}+\sqrt{2}}{4} & \frac{\sqrt{2}+1}{4} & \frac{-\sqrt{2}+1}{4} \\ \frac{-\sqrt{3}+\sqrt{2}}{4} & \frac{-\sqrt{3}-\sqrt{2}}{4} & \frac{-\sqrt{2}+1}{4} & \frac{\sqrt{2}+1}{4} \end{bmatrix}$$

é a matriz da rotação ρ na base canônica.

4. Busca na rede

O PAGERANK foi o algoritmo originalmente usado pelo Google para ordenar as páginas obtidas em uma busca. Como uma última aplicação da álgebra linear, veremos como o PAGERANK usa um problema de autovalores e autovetores para determinar o peso que especifica a posição de uma dada página em uma busca.

4.1. Grafos na web. Há muitas maneiras de interpretar a web como um grafo.

A ser escrito!

Apesar de ser estocástica, a matriz A pode ter coeficientes nulos, o que permite que o autoespaço associado ao autovalor 1 tenha dimensão maior que um. Para contornar este problema, substituímos A por uma outra matriz, que chamaremos de G . Se S for a matriz $n \times n$ cujas entradas são todas iguais a $1/n$, definimos

$$G = (1 - m)A + mS,$$

em que m é um número real entre 0 e 1. A versão original do PAGERANK utilizava $m = 0,15$. O ponto crucial é que esta matriz deve ser estocástica por coluna e ter todas as suas entradas positivas. Mas A e S são estocásticas por coluna, o que equivale a dizer que

$$u_0^t A = u_0^t S = u_0^t, \text{ em que } u_0^t = (1, 1, \dots, 1);$$

onde

$$u_0^t G = (1-m)u_0^t A + mu_0^t S = ((1-m)+m)u_0^t = u_0^t,$$

comprovando que G também é estocástica por coluna. Por outro lado, como a soma de matrizes é feita entrada a entrada, é suficiente discutir o que acontece com cada entrada de G para entender porque esta matriz é positiva. Mas

$$G(i, j) = (1-m)A(i, j) + m\frac{1}{n}.$$

Há dois casos a considerar. Se $A(i, j) = 0$, então

$$G(i, j) = m\frac{1}{n}$$

é claramente positivo; mas se $A(i, j) > 0$ e

$$\mu = \min\{A(i, j), 1/n\},$$

então

$$G(i, j) = (1-m)A(i, j) + m\frac{1}{n} \geq (1-m)\mu + m\mu = \mu$$

que tem que ser um número positivo. Portanto, em qualquer dos casos, $G(i, j) > 0$ e G é uma matriz positiva, como havíamos afirmado. A razão pela qual dispomos este esforço adicional para obter uma matriz positiva vai se tornar clara nos próximos artigos.

4.2. Matrizes estocásticas por colunas. No artigo anterior vimos como representar o grafo dirigido resultante de uma busca na forma de uma matriz estocástica por coluna positiva G . Como vimos no artigo 1.2 que matrizes estocásticas por colunas sempre têm 1 como um de seus autovalores. Nesta seção provaremos que se uma matriz estocástica por coluna também é positiva, então o autoespaço de 1 (1) *contém um autovetor p cujas coordenadas são todas positivas e (2) tem dimensão igual a um*.

As entradas de p correspondem aos pesos de cada uma das páginas pesquisadas. Feita a busca e calculado p , as páginas são listadas em ordem decrescente de peso. A demonstração de (1) e (2) será feita em etapas. Começamos introduzindo uma maneira de medir o comprimento de um vetor que é mais adequada às matrizes estocásticas por colunas do que a norma euclidiana usual.

Se $v = (v_1, \dots, v_n)$. Definimos sua 1-norma como sendo o número

$$|v|_1 = |v_1| + \dots + |v_n|.$$

É fácil mostrar que a 1-norma satisfaz às propriedades esperadas de uma norma; isto é:

- $|v + w|_1 \leq |v|_1 + |w|_1$;
- $|\lambda v|_1 \leq |\lambda| |v|_1$;
- $|v|_1 \geq 0$;
- $|v|_1 = 0$ se, e somente se, $v = 0$.

Por exemplo, se

$$v = [v_1, \dots, v_n]^t \quad \text{e} \quad w = [w_1, \dots, w_n]^t$$

então, pela definição da 1-norma

$$(159) \quad |v + w|_1 = |v_1 + w_1| + \dots + |v_n + w_n|.$$

Contudo, o módulo de um número real satisfaz a desigualdade triangular, donde

$$|v_i + w_i| \leq |v_i| + |w_i| \quad \text{para todo } 1 \leq i \leq n.$$

Substituindo estas desigualdades em (159), obtemos

$$|v + w|_1 \leq |v_1| + |w_1| + \dots + |v_n| + |w_n|;$$

que, reagrupando as parcelas, nos dá a desigualdade desejada. As outras propriedades ficam como exercício para você.

 Note que, embora possa ocorrer que a 1-norma de um vetor seja igual à sua norma euclidiana, em geral as duas são muito diferentes; por exemplo, as normas do vetor

$$w = (1, 2, 3, \dots, n) = \sum_{i=1}^n ie_i$$

são

$$\|w\| = \sqrt{\frac{2n^3 + 3n^2 + n}{6}} \quad \text{e} \quad |w|_1 = \frac{n(n+1)}{2}.$$

É razoável esperar que a 1-norma seja mais adequada ao estudo das matrizes estocásticas por colunas do que a norma euclidiana, uma vez que uma matriz estocástica por coluna A tem que satisfazer $|A(:, i)| = 1$, para todo $1 \leq i \leq n$. Na verdade, a 1-norma admite uma tradução matricial muito conveniente para o que faremos adiante. Denotando por $|v|$ o vetor

$$(|v_1|, \dots, |v_n|)$$

cujas coordenadas são os módulos das entradas de v e por u_0 o vetor

$$(1, \dots, 1) = \sum_{i=1}^n e_i$$

podemos escrever

$$|v|_1 = u_0^t |v|.$$

Com isto podemos voltar à discussão das questões (1) e (2) enunciadas acima.

Por todo o resto deste artigo suporemos que G é uma matriz estocástica por coluna positiva $n \times n$. Começaremos provando uma propriedade referente a vetores que têm coordenadas positivas e negativas, e que, para os propósitos deste curso, chamaremos de *mistas*.

Propriedade 1: V_1 não contém vetores mistos.

Suponha, por contradição, que v seja um vetor misto em V_1 . Se $v = (v_1, \dots, v_n)$, então a igualdade $Gv = v$ pode ser reescrita na forma

$$v_i = G_i v;$$

em que G_1, \dots, G_n são as *linhas* de G . Como G é positiva e v é misto, temos que

$$|v_i| = |G_i v| < G_i |v|.$$

Como

$$\sum_{i=1}^n G_i |v| = G |v|$$

podemos concluir de $u_0^t G = u_0^t$, que

$$|v|_1 = u_0^t |v| < u_0^t G |v| = |v|_1$$

que nos dá a esperada contradição. Note que o que provamos é, na verdade, mais forte que o enunciado original de (1).

Passando a (2), devemos provar

Propriedade 2: $\dim(V_1) = 1$.

Mais uma vez o argumento será por contradição. Digamos que v e v' sejam vetores linearmente independentes em V_1 . Se

$$d = u_0^t v = \sum_{i=1}^n v_i \quad \text{e} \quad d' = u_0^t v' = \sum_{i=1}^n v'_i$$

então o vetor

$$w = d'v - dv' = (u_0^t v')v - (u_0^t v)v' \neq 0,$$

satisfaz

$$u_0^t w = (u_0^t v')(u_0^t v) - (u_0^t v)(u_0^t v') = 0;$$

de forma que w tem que ser misto, o que viola a propriedade 1, provando o que desejávamos. Combinando as propriedades 1 e 2, mostramos a seguinte proposição.

PROPOSIÇÃO 4.1. *Se G é estocástica por coluna positiva então existe um vetor unitário \mathbf{p} de coordenadas positivas que gera o autoespaço associado a 1.*

Como já mencionamos, este vetor especial, será considerado como um vetor de pesos. As páginas de maior peso serão listadas no início, quando a busca for apresentada ao usuário.

4.3. Calculando o vetor peso. Para que esta maneira de ponderar as páginas listadas em uma busca seja viável, devemos ser capazes de determinar o vetor peso de forma altamente eficiente. Afinal uma busca típica relaciona milhões de resultados e não toma mais que uma fração de segundos. O seguinte resultado será necessário à justificativa do funcionamento do algoritmo. Denotaremos por u_0 o vetor $(1, \dots, 1)$, e por W o subespaço

$$W = \{v \in \mathbb{R}^n \mid u_0^t v = 0\}.$$

Portanto, na terminologia do artigo anterior, todos os vetores não nulos de W são mistos.

PROPOSIÇÃO 4.2. *Seja G uma matriz estocástica por coluna positiva. Se $w \in W$ então*

- (a) $Gw \in W$;
- (b) $|Gw|_1 \leq c|w|_1$;

para alguma constante $0 < c < 1$.

Como $u_0^t G = u_0^t$, temos que

$$u_0^t Gw = u_0^t w = 0,$$

para todo $w \in W$, o que prova (a). Antes de passar à demonstração de (b), vamos expressar a 1-norma de w como um produto de matrizes. Como w é misto, não podemos usar u_0 , como fizemos no artigo anterior. A solução é definir o vetor de sinais $s(w)$ cujas coordenadas (s_1, \dots, s_n) satisfazem

$$s_i = \begin{cases} 1 & \text{se } w_i > 0 \\ -1 & \text{se } w_i < 0 \\ 0 & \text{se } w_i = 0. \end{cases}$$

Usando este vetor, podemos escrever

$$|w|_1 = s(w)^t w.$$

Em particular,

$$|Gw|_1 = s(Gw)^t Gw.$$

Contudo, se G_j for a j -ésima coluna de G , temos que

$$-u_0^t G_j < s(Gw)^t G_j < u_0^t G_j$$

pois as entradas de G_j são todas positivas. Como G é estocástica por coluna ,

$$-1 < s(Gw)^t G_j < 1.$$

o que nos permite concluir que

$$c = \max\{|s(Gw)^t G_j| \mid 1 \leq j \leq n\} < 1.$$

Logo, de

$$|Gw|_1 = s(Gw)^t Gw = \sum_{j=1}^n (s(Gw)^t G_j) w_j$$

segue que

$$|Gw|_1 = s(Gw)^t Gw \leq c \sum_{j=1}^n w_j \leq c|w|_1;$$

que nos dá a desigualdade desejada.

Com isto podemos enunciar e provar um algoritmo que, tendo como entrada uma matriz estocástica por coluna positiva G e uma tolerância $\epsilon > 0$, calcula um autovetor de G associado ao autovalor 1:

- Inicializa:** escolha um vetor unitário positivo $v^{(0)}$ qualquer;
- Iteração:** enquanto $|v^{k+1} - v^{(k)}|_1 > \epsilon$ repita $v^{k+1} = Gv^{(k)}$;
- Saída:** retorne $\mathbf{p} = v^{(k)}$.

Para provar que o algoritmo funciona, devemos mostrar

$$\lim_{k \rightarrow \infty} v^{(k)}$$

é um autovetor de G associado a 1. Seja \mathbf{p} o autovetor positivo de 1-norma igual a 1 associado a 1. Digamos que $\delta = \mathbf{p} - v^{(0)}$. Se todas as entradas de δ fossem positivas, teríamos que cada uma das coordenadas de \mathbf{p} seria maior que sua correspondente em $v^{(0)}$. Como $v^{(0)}$ e \mathbf{p} são ambos vetores positivos, isto só pode ocorrer se $|\mathbf{p}|_1 > |v^{(0)}|_1$, o que não é possível, já que ambos têm 1-norma igual a 1. Como o mesmo argumento funciona se todas as coordenadas de δ forem negativas, δ tem que ser um vetor misto. Aplicando a proposição,

$$|G\delta|_1 < c|\delta|_1;$$

onde segue por indução em k que

$$|G^k \delta|_1 < c^k |\delta|_1.$$

Como δ está fixo, temos que

$$\lim_{k \rightarrow \infty} G^k \delta = 0,$$

pois o único vetor cuja 1-norma é zero é o vetor nulo. Contudo,

$$\lim_{k \rightarrow \infty} G^k \delta = q - \lim_{k \rightarrow \infty} v^{(k)};$$

de modo que $\mathbf{p} = \lim_{k \rightarrow \infty} v^{(k)}$, é um autovetor positivo de G associado a 1.

Por exemplo, a matriz

$$A = \begin{bmatrix} 0.5168758 & 0.3896685 & 0. & 0.2140355 & 0.1937507 \\ 0.4831242 & 0.6103315 & 0. & 0.3231644 & 0.1426122 \\ 0. & 0. & 1. & 0.1735320 & 0.2065034 \\ 0. & 0. & 0. & 0.2610884 & 0.2998340 \\ 0. & 0. & 0. & 0.0281797 & 0.1572998 \end{bmatrix}$$

é estocástica *por colunas*, mas não é positiva, de modo que, tomando $m = 0.15$, temos que

$$G = (1 - m)A + mS = \begin{bmatrix} 0.4693445 & 0.3612182 & 0.03 & 0.2119302 & 0.1946881 \\ 0.4406555 & 0.5487818 & 0.03 & 0.3046898 & 0.1512203 \\ 0.03 & 0.03 & 0.88 & 0.1775022 & 0.2055279 \\ 0.03 & 0.03 & 0.03 & 0.2519251 & 0.2848589 \\ 0.03 & 0.03 & 0.03 & 0.0539528 & 0.1637048 \end{bmatrix}.$$

Aplicando o algoritmo a G com $e = 10^{-5}$, obtemos, ao cabo de dez iterações, o vetor

$$v^{(10)} = [0.2815365, 0.3404141, 0.2916703, 0.0503566, 0.0360227]^t.$$

Os erros cometidos a cada iteração são listados na tabela 1. Fica claro mesmo neste exemplo simples que, a despeito de sua incrível eficiência, não se trata de um algoritmo fácil de executar usando apenas lápis e papel.

Iteração	1	2	3	4	5
Erro	0.4262234	0.1443139	0.0412772	0.0112782	0.0030381
Iteração	6	7	8	9	10
Erro	0.0008146	0.0002181	0.0000584	0.0000156	0.0000042

TABELA 1. Erros por iteração

Exercícios

1. Ache os autovalores e autovetores das matrizes:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & -3 & 4 \\ 0 & 3 & 5 \\ 0 & 0 & -1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 2 \\ -1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix},$$

$$D = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix}, \quad E = \begin{bmatrix} 1 & 3 & -2 \\ 2 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad F = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 0 \\ 3 & 3 & 1 \end{bmatrix}.$$

2. Determine a base dos autoespaços de cada uma das matrizes do exercício anterior.
3. Sejam $v_1 = (1, 0, 1)$, $v_2 = (1, -1, 0)$ e $v_3 = (0, 0, 1)$.
- Mostre que estes vetores formam uma base B do \mathbb{R}^3 .
 - Determine a matriz na base B do operador linear definido por
- $$T(v_1) = 2v_1, \quad T(v_2) = -v_2 \text{ e } T(v_3) = -v_2.$$
- Determine a matriz na base canônica deste mesmo operador.
4. Determine a matriz na base canônica de um operador linear do \mathbb{R}^3 que tenha autovalores $1, -2$ e 0 associados aos autovetores $(1, 0, -1)$, $(1, 0, 0)$ e $(0, 1, 1)$.
5. Seja T o operador linear do \mathbb{R}^3 definido por $T(x, y, z) = (x + y + z, 2y + z, 2y + 3z)$.
- Ache os autovalores e autovetores de T .
 - Ache uma base para cada autoespaço de T .
 - Este operador é diagonalizável? Se a resposta for sim, determine uma base B de autovetores para T e calcule a matriz $(T)_B$.
6. Seja T o operador linear do \mathbb{R}^3 definido por $T(x, y, z) = (2x + y, y - z, 2y + 4z)$.
- Ache os autovalores e autovetores de T .
 - Ache uma base para cada autoespaço de T .
 - Este operador é diagonalizável? Se a resposta for sim, determine uma base B de autovetores para T e calcule a matriz $(T)_B$.
7. Quais das matrizes A abaixo são diagonalizáveis? Para aquelas que forem diagonalizáveis determine uma matriz M inversível M tal que $M^{-1}AM$ é diagonal.

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

8. Calcule A^{100} , quando A for a matriz

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

9. Verifique se as matrizes

$$\begin{bmatrix} 0 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1/3 & 2/3 & -2/3 \\ 2/3 & 1/3 & 2/3 \\ 2/3 & -2/3 & -1/3 \end{bmatrix} \text{ e } \begin{bmatrix} \sqrt{2}/2 & -1/2 & -1/2 \\ 0 & \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & 1/2 & 1/2 \end{bmatrix}$$

representam rotações. Em caso afirmativo calcule o eixo e o ângulo de rotação.

10. Seja R uma rotação de eixo ℓ em \mathbb{R}^3 e $v = (1, 1, 1)$ um vetor ortogonal a ℓ . Sabendo-se que $Rv = (1, -1, 1)$, determine:
- o cosseno do ângulo de rotação de R ;
 - o eixo da rotação R ;
 - a matriz de R na base canônica.
11. Determine números reais a , b e c de modo que a matriz

$$Q = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{14} & a \\ 1/\sqrt{3} & 2/\sqrt{14} & b \\ 1/\sqrt{3} & -3/\sqrt{14} & c \end{bmatrix}$$

seja uma rotação do \mathbb{R}^3 .

12. Considere a matriz

$$Q = \begin{bmatrix} 1/3 & 2/3 & -2/3 \\ a & 2/3 & 1/3 \\ b & 1/3 & c \end{bmatrix}.$$

Determine valores para a , b e c de forma que Q descreva uma rotação de \mathbb{R}^3 . Ache o eixo e o cosseno do ângulo de rotação de Q .

13. Quais das seguintes matrizes representam rotações do \mathbb{R}^4 ? Para cada uma delas, determine planos invariantes ortogonais e o ângulo de rotação em cada um destes planos.

$$(a) \quad \frac{1}{7} \begin{bmatrix} -2 & 5 & -2 & -4 \\ 5 & -2 & -2 & -4 \\ -2 & -2 & 5 & -4 \\ -4 & -4 & -4 & -1 \end{bmatrix} \quad (b) \quad \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (c) \quad \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix}$$

14. Seja T o operador de \mathbb{R}^3 definido por $T(x, y, z) = (x + 4y + 2z, 4x - 5y - 4z, 2x - 4y + z)$.
- Ache uma base ortonormal B do \mathbb{R}^3 constituída por autovetores de T .
 - Determine uma matriz M tal que $M^t(T)_\varepsilon M = (T)_B$.
15. Seja T o operador de \mathbb{R}^3 definido por $T(x, y, z) = (2x + y, x + y + z, y - 3z)$. Mostre que T é autoadjunto e determine uma base de \mathbb{R}^3 formada por autovetores de T .
16. Considere o operador do \mathbb{R}^3 cuja matriz na base canônica é

$$\begin{bmatrix} \frac{14}{3} & 0 & -\frac{4}{3} \\ 0 & \frac{10}{3} & -\frac{4}{3} \\ -\frac{4}{3} & -\frac{4}{3} & 4 \end{bmatrix}$$

Encontre os vértices de um paralelepípedo que tenha a origem como um de seus vértices e que seja levado por T em um cubo de aresta igual a 70 unidades.

17. Sejam T um operador autoadjunto, u um autovetor de T e w um vetor perpendicular a u . Prove que Tw também é perpendicular a u .

18. Considere o operador linear T de \mathbb{R}^3 definido por

$$T(x, y, z) = (y + z, x + z, x + y).$$

Determine:

- (a) os autovalores de T ;
- (b) os autoespaços de T ;
- (c) uma base β de autovetores de T ;
- (d) a matriz de mudança de base de β para a base canônica do \mathbb{R}^3 .

19. Seja S o plano do \mathbb{R}^4 gerado pelos vetores $(1, 1, 0, 0)$ e $(1, 0, 1, 1)$. Determine

- (a) o complemento ortogonal S^\perp de S ;
- (b) um operador linear T de \mathbb{R}^4 cujo núcleo é S e cuja imagem é S^\perp em S .

20. Seja U o plano de equação $x - y + 2z = 0$ e ℓ a reta gerada por $(1, 1, 2)$.

- (a) Determine um operador linear de \mathbb{R}^3 cujo núcleo é U e cuja imagem é ℓ .
- (b) Prove que um operador que satisfaz as propriedades de (a) não pode ser autoadjunto.

21. Determine *todos* os valores possíveis de a , b e c para os quais a matriz

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

corresponda a um operador diagonalizável.

22. Ache um paralelepípedo que seja levado em um cubo de lado 8 pelo operador linear T de \mathbb{R}^3 definido por

$$T(x, y, z) = (2x + y + z, x + 2y + z, x + y + 2z).$$

23. Segundo o *Teorema do Valor Intermediário*, se uma função contínua $f : \mathbb{R} \rightarrow \mathbb{R}$ satisfaça $f(a) < 0$ e $f(b) > 0$ para dois números reais $a < b$, então existe um número real r entre a e b no qual f se anula. Use isto para mostrar que todo polinômio de grau ímpar tem uma raiz real.

24. Mostre que se um polinômio cujos coeficientes são reais tem um número complexo α como uma de suas raízes, então o conjugado de α também é raiz do mesmo polinômio.

25. Um polinômio de grau n cujos coeficientes são reais é *irreduzível* se não pode ser escrito como o produto de dois polinômios, ambos de grau estritamente menor que n .
- Mostre que nenhum polinômio de grau ímpar e coeficientes reais é irreduzível.
 - Determine uma condição sobre $a, b, c \in \mathbb{R}$ para que o polinômio $ax^2 + bx + c$ seja irreduzível.
26. Segundo o *Teorema Fundamental da Álgebra* um polinômio $f(x)$, de grau n e coeficientes reais, pode ser fatorado na forma

$$f(x) = a \prod_{i=1}^n (x - \alpha_i)$$

em que $a \in \mathbb{R}$ e os α 's são números complexos. Use este teorema e os exercícios anteriores para mostrar que f só pode ser irreduzível se $n = 1$ ou $n = 2$.

Referências Bibliográficas

- [1] M. Born, *My life: recollections of a Nobel laureate*, Charles Scribner's Sons (1978).
- [2] A. J. Crilly, *Arthur Cayley: mathematician laureate of the Victorian Age*, Johns Hopkins University Press (2006).
- [3] F. Klein, *Development of mathematics in the 19th century*, translated by M. Ackerman with an appendix by R. Hermann, Math Sci Press (1979).
- [4] A. S. Householder, *Unitary triangularization of a nonsymmetric matrix*, J. Assoc. Comput. Mach. **5** (1958) 339–342.
- [5] B. Palais, R. Palais e S. Rodi, *A Disorienting Look at Euler's Theorem on the Axis of a Rotation*, The American Mathematical Monthly, **116** (2009), 892–909.
- [6] P. C. Shields, *The Spectral Theorem for Real Symmetric Matrices*, The American Mathematical Monthly, **77** (1970), 862–863.
- [7] L. N. Trefethen e D. Bau, *Numerical linear algebra*, SIAM (1997).