Álgebra Linear

Um segundo curso

Hamilton Prado Bueno

Universidade Federal de Minas Gerais
Departamento de Matemática

Georg, Claudia und Miriam Müller
gewidmet

Prefácio

Esse texto é uma adaptação de parte de um livro de Álgebra Linear que considero uma obra prima: o livro "Linear Algebra", de P. Lax.

Adaptar o texto de P. Lax é, no fundo, uma temeridade. Não acredito que aquele texto possa ser melhorado. Por outro lado, ele foi escrito tendo como propósito um curso de pós-graduação no Courant Institute. É um texto denso e sintético. Após umas poucas aulas cheguei à conclusão que os meus alunos dificilmente conseguiriam acompanhá-lo. Daí surgiu a necessidade dessa adaptação. Tentei esmiuçar algumas passagens; substituí demonstrações elegantes, mas sintéticas, por outras mais diretas. Tentando poupar algum tempo na exposição de assuntos, suprimi material que servia de motivação. Incluí pré-requisitos e reordenei parte do material exposto. Aumentei a ênfase em espaços vetoriais reais. Mas, tendo concluído a adaptação de quase todos os oito primeiros capítulos (de um total de dezessete, mais oito apêndices, tudo isso em apenas 250 páginas!), a comparação do texto original com a adaptação é apenas um desprestígio para o primeiro. Mais do que isso, com o decorrer do curso, verifiquei que algumas passagens que os alunos julgavam incompreensíveis no livro de Lax puderam ser absorvidas. Ou seja, bastou um pouco de maturidade matemática para tornar aquele texto inteligível.

O presente texto é dirigido a alunos que cursam um segundo curso de Álgebra Linear. Na preparação dessa adaptação fiz uso, principalmente, do texto "Geometria Analítica e Álgebra Linear", 2a. parte, do Prof. Reginaldo J. Santos [15]. Esse é bastante direto, apresentando a Álgebra Linear de um ponto de vista bastante adequado à sua utilização por engenheiros e não-matemáticos. É difícil encontrar um livro tão bem escrito de introdução à Álgebra Linear. Sugeri esse texto como leitura complementar aos meus alunos, principalmente àqueles que haviam cursado seu primeiro curso de Álgebra Linear há algum tempo. Ele apresenta demonstrações simples de resultados que, em outros textos, tem tratamento muito mais complicado: compare-se, por exemplo, as demonstrações do Teorema de Cayley-Hamilton daquele texto (aqui transcrita) e aquela do livro do Prof. Elon Lima. Na escolha de material complementando o livro de Lax, utilizei as notas de aula do Prof. Marcos Montenegro [13] e o livro de Leon [11]. O primeiro foi especialmente útil no tratamento de espaços vetoriais reais e o segundo na apresentação de alguns resultados da Álgebra Linear Numérica.

Os capítulos desse texto cobrem um curso de Álgebra Linear usual: espaços vetoriais e bases, o espaço dual, aplicações lineares e matrizes, determinantes, o Teorema da Decomposição Primária e a Forma Canônica de Jordan, espaços euclidianos, formas quadráticas, o Teorema Espectral para operadores normais (e, com isso, operadores unitários e ortogonais) e, finalmente, o Teorema de Valores Singulares.

Faço alguns comentários sobre os capítulos desse texto. Observo que, apesar de todas as noções básicas da Álgebra Linear serem apresentadas, alguns capítulos foram escritos no espírito "revisão", notadamente os capítulos 1, 3 e parte do capítulo 6. Assim, é pressuposto que o aluno tenha alguma familiaridade com matrizes e sistemas lineares, bases e o produto interno no espaço $\mathbb{R}^n$.

O capítulo 1 introduz espaços vetoriais e bases. Os espaços vetoriais são considerados apenas sobre os corpos $\mathbb{R}$ ou $\mathbb{C}$, o que é coerente com a linha geral do texto, que é voltado para a área de Análise. Os alunos que assistiram o curso não possuíam formação em Álgebra. Isso tornou necessária uma apresentação detalhada do espaço quociente. Incluí no texto alguns dos exercícios que procuravam esclarecer o assunto, mas não a interpretação geométrica apresentada em sala de aula. Apesar disso, é bom salientar que o espaço quociente é usado apenas duas vezes: uma na demonstração do Teorema do Núcleo e da Imagem (que também possui uma prova alternativa, sem o uso desse conceito) e outra na demonstração da Forma Canônica de Jordan, quando apenas é necessária a idéia do que é uma base do espaço quociente e não do espaço propriamente dito. Não é difícil adaptar aquela demonstração sem se mencionar o espaço quociente, de modo que sua apresentação fica a critério do instrutor. Por outro lado, a introdução do espaço quociente na demonstração do Teorema do Núcleo e da Imagem unifica conceitos: a mesma demonstração se repete no estudo de outras estruturas algébricas. Apresentei o capítulo 1 em ritmo acelerado, já que seu conteúdo era familiar aos alunos do curso.

O capítulo 2 trata do espaço dual e apresenta uma primeira versão do Teorema de Representação de Riesz (para espaços de dimensão finita). Geralmente o dual e o bidual são apresentados após a introdução de espaços de aplicações lineares, como casos particulares desses. O texto inverte essa ordem para dar exemplo de um isomorfismo canônico entre espaços vetoriais. Com modificações corriqueiras no capítulo 3, o instrutor pode optar por não apresentar esse capítulo.

O capítulo 3 começa por mostrar que a definição de multiplicação de matrizes é uma conseqüência natural da composição de aplicações lineares. Nesse capítulo também são tratados outros tópicos fundamentais de um curso de Álgebra Linear: matrizes e representações de aplicações lineares, núcleo e imagem de uma aplicação linear, sistemas lineares, espaço-linha e espaço-coluna, etc. Sua apresentação foi rápida; decidi não expor a sua última seção.

O capítulo 4 apresenta determinantes, desde o ponto de vista de permutações. Procurei evitar uma apresentação demasiadamente abstrata. Incluí material sobre ciclos e transposições que não é estritamente necessário ao estudo de determinantes[1]; além disso, adeqüei o ritmo da minha exposição à pouca familiaridade dos alunos com esses conceitos. Ainda assim, esses quatro primeiros capítulos foram cobertos em aproximadamente 30 horas de aula de um curso semestral de 90 horas.

O capítulo 5 apresenta o Teorema de Cayley-Hamilton e as formas canônicas fundamentais: o Teorema da Decomposição Primária e a Forma Canônica de Jordan. A primeira seção do capítulo é escrita de forma a apresentar os resultados básicos sobre diagonalização de matrizes. Então se estudam os polinômios matriciais e se demonstra o Teorema de Cayley-Hamilton. As provas dos Teoremas da Decomposição Primária (que Lax denomina, no caso de espaços vetoriais sobre $\mathbb{C}$, de Teorema Espectral) e da Forma Canônica de Jordan são bastante objetivas, e se apóiam em resultados que estão explicitamente demonstrados no texto. Decidi apresentar a versão real desses dois teoremas, o que não constava do texto original. Vários exemplos são dirigidos à Forma Canônica de Jordan. Dediquei aproximadamente 25 horas de aula a esse capítulo e, no decorrer de sua exposição, voltei repetidamente à demonstração do Teorema da Decomposição

[1]Interpretando adequadamente, a apresentação de Lax sobre o sinal de uma permutação é mais concisa. Lax deixa como exercício a demonstração de que uma permutação é um produto de transposições.

Primária. Achei proveitoso esse procedimento: as idéias fundamentais desse teorema, bem como seu extraordinário significado, ficam melhor compreendidos se sua importância é constantemente salientada.

O capítulo seguinte trata de espaços com produto interno. Lax segue a tradição bourbakista de apresentá-los apenas após o estudo de espaços vetoriais gerais. Mantive esse ordenamento, apesar de achá-lo demasiadamente purista para os meus propósitos, que eram enfatizar espaços de dimensão finita. O capítulo é leve e pode ser exposto mais rapidamente, mesmo assim trazendo algum alívio aos alunos após a maratona do capítulo anterior, já que apresenta tópicos familiares de um primeiro curso de Álgebra Linear. (Mesmo assim, acho que o instrutor deve ressaltar o aspecto geométrico introduzido conjuntamente com o produto interno. Por exemplo, o processo de ortogonalização de Gram-Schmidt pode ser justificado em casos bi- e tridimensionais. Mais do que isso, no caso de espaços de dimensão n, uma representação decompondo-o em um eixo vertical e seu complementar ortogonal é adequada: muitas demonstrações podem ser, assim, geometricamente justificadas). Em coerência com o caminho voltado para a Análise, algumas propriedades da norma de uma aplicação linear são apresentadas. Também são estudadas as relações entre o núcleo e a imagem de uma aplicação linear e de sua adjunta, bem como algumas propriedades básicas de isometrias.

Voltando a diminuir o ritmo da exposição em sala de aula, o capítulo 7 trata das principais formas canônicas em espaço com produto interno: o Teorema Espectral para operadores normais, operadores unitários e ortogonais. O capítulo começa tratando do Teorema de Sylvester e então apresenta, como um refinamento, a diagonalização de matrizes simétricas, cuja demonstração é feita à partir do Teorema da Decomposição Primária. Esse enfoque unifica conceitos que usualmente são apresentados separadamente: formas bilineares simétricas e diagonalização de matrizes simétricas. As versões reais dos teoremas também estão presentes, diferindo mais uma vez do texto original. Os exercícios do capítulo procuram esclarecer duas relações de equivalência: a semelhança de matrizes ($B = P^{-1}AP$) e a equivalência de matrizes ($B = P^TAP$). Dediquei aproximadamente 20 horas de aula a esse capítulo.

O capítulo 8, que não consta no livro de Lax e não foi apresentado no curso, trata de decomposições matriciais: LU, Cholesky, Schur, QR e valores singulares, resultados especialmente úteis na Álgebra Linear Numérica. Decidi incluí-lo por dois motivos. Em primeiro lugar, alguns desses tópicos (a saber, as decomposições LU, QR e em valores singulares) são apenas a formulação matricial de resultados conhecidos. Já a decomposição de Schur possibilita uma demonstração independente do teorema de diagonalização de operadores normais, enquanto Cholesky desvela o vínculo entre a decomposição LU e matrizes positivas-definidas. Mas, mais importante do que isso, esses temas usualmente não são abordados em apresentações tradicionais, e isso significa ignorar todo o desenvolvimento proporcionado pela introdução de métodos numéricos no estudo da Álgebra Linear. O capítulo pode ser apresentado em combinação com capítulos anteriores, sem um acréscimo substancial em termos de tempo de aula.

Os exercícios incluídos no livro, alguns formulados por mim mesmo e outros compilados de diversos textos, têm vários graus de dificuldade. Algumas vezes, eles introduzem notações e conceitos que serão usados livremente no resto do texto. Alguns indicam demonstrações alternativas de resultados expostos. Outros complementam o material apresentado, sugerindo generalizações.

Belo Horizonte, fevereiro de 2002

Hamilton Prado Bueno

Sumário

Capítulo 1

Base e Dimensão

1.1 Espaços vetoriais

Denotaremos por $\mathbb{K}$ o corpo $\mathbb{R}$ ou o corpo $\mathbb{C}$.

Definição 1.1.1 *Um* **espaço vetorial** X *sobre o corpo* $\mathbb{K}$ *é um conjunto cujos elementos* (*chamados* **vetores**) *podem ser somados e multiplicados por* **escalares**, *isto é, os elementos do corpo* $\mathbb{K}$. *Se* $x, y, z \in X$ *e* $\lambda, \mu \in \mathbb{K}$, *as seguintes propriedades devem ser satisfeitas pela adição e multiplicação por escalar:*

(*i*) $x + y \in X$ (*fechamento*);

(*ii*) $(x + y) + z = x + (y + z)$ (*associatividade*);

(*iii*) $x + y = y + x$ (*comutatividade*);

(*iv*) *existe* $0 \in X$ *tal que* $x + 0 = x$ (*elemento neutro*);

(*v*) *existe* $(-x) \in X$ *tal que* $x + (-x) = 0$ (*inverso aditivo*);

(*vi*) $\lambda x \in X$ (*fechamento*);

(*vii*) $\mu(\lambda x) = (\mu\lambda)x$ (*associatividade*);

(*viii*) $\lambda(x + y) = \lambda x + \lambda y$ (*distributividade*);

(*ix*) $(\lambda + \mu)x = \lambda x + \mu x$ (*distributividade*);

(*x*) $1x = x$ (*regra da unidade*).

Denotaremos $x + (-y)$ simplesmente por $x - y$ (veja exercício 1).

Exemplo 1.1.2 O conjunto $\mathbb{K}^n = \{(x_1, x_2, \ldots, x_n) : x_i \in \mathbb{K}\ (i = 1, \ldots, n)\}$ com a adição e multiplicação por escalar definidas coordenada a coordenada é um espaço vetorial. O conjunto $\mathcal{F}$ de todas as funções $\{f : S \to \mathbb{K}\}$ definidas num conjunto arbitrário S e com as operações de adição e multiplicação por escalar usualmente definidas é também um espaço vetorial. O mesmo acontece com o conjunto $\mathcal{P}$ de todos os polinômios com coeficientes em $\mathbb{K}$ ou o subconjunto $\mathcal{P}_n$ de todos os polinômios de grau menor do que n. ◀

Definição 1.1.3 *Um subconjunto Y de um espaço vetorial X é um* **subespaço** *se seus elementos satisfazem às propriedades que definem o espaço vetorial X.*

Exemplo 1.1.4 O subconjunto de $\mathbb{K}^n$ de todos os vetores cuja primeira coordenada é nula é um subespaço de $\mathbb{K}^n$. Se $S = \mathbb{R}$, os subconjunto de $\mathcal{F}$ formado por todas as funções contínuas ou por todas as funções de período π são subespaços de $\mathcal{F}$. O mesmo acontece com o subconjunto de $\mathcal{P}$ formado por todos os polinômios de grau par. ◀

Definição 1.1.5 *Sejam X e Y espaços vetoriais sobre o corpo $\mathbb{K}$. Uma aplicação*

$$T : X \to Y$$

satisfazendo

$$T(x + \lambda y) = Tx + \lambda Ty$$

para quaisquer $x, y \in X$ e $\lambda \in \mathbb{K}$ é chamada **transformação linear** *ou* **aplicação linear**. *Se $X = Y$ também chamamos T de* **operador linear**.

Se T é uma bijeção, dizemos que T é um **isomorfismo** *e que os espaços X e Y são* **isomorfos**.

Observação 1.1.6 Note que, na definição de aplicação linear, estamos denotando as operações nos espaços vetoriais X e Y da mesma maneira: em $T(x + \lambda y)$, a soma $x + \lambda y$ ocorre no espaço X, enquanto em $Tx + \lambda Ty$ ela ocorre em Y. ◀

1.2 Somas diretas

Definição 1.2.1 *Sejam A, B subconjuntos de um espaço vetorial X. Denotamos $A + B$ o conjunto de todos os vetores $x + y$, com $x \in A$ e $y \in B$.*

Proposição 1.2.2 *Sejam U, V subespaços de X. Então $U + V$ é subespaço de X. O subespaço $U + V$ é chamado* **soma** *dos subespaços U e V.*

Demonstração: Se $z_1 = x_1 + y_1$ e $z_2 = x_2 + y_2$ são elementos de $U + V$ e $\lambda \in \mathbb{K}$, então claramente $\lambda z_1 + z_2 \in U + V$ (veja exercício 3).

□

Definição 1.2.3 *Sejam U, V subespaços de X. O subespaço $W = U + V$ é a* **soma direta** *dos subespaços U e V se cada elemento de $w \in W$ pode ser escrito de maneira única como*

$$w = x + y.$$

Nesse caso denotamos $W = U \oplus V$.

A definição de soma direta pode ser generalizada para a soma de um número finito de subespaços de X.

Proposição 1.2.4 *O subespaço $W = U + V$ é a soma direta dos subespaços U, V de X se, e somente se, $U \cap V = \{0\}$.*

Demonstração: Suponhamos que $W = U \oplus V$. Se $z \in U \cap V$ então $w = x + y$ também pode ser escrito como $w = (x + z) + (y - z)$. Como a decomposição $w = x + y$ é única, devemos ter $x = x + z$ e $y = y - z$. Assim, $z = 0$ (veja exercício 2).

Reciprocamente, suponhamos que $x_1 + y_1$ e $x_2 + y_2$ sejam duas decomposições de $w \in W$. Então $x_1 - x_2 = y_2 - y_1$ pertencem simultaneamente a U e V. Logo $x_1 - x_2 = 0 = y_2 - y_1$, garantindo a unicidade da decomposição.

□

1.3 Bases

Definição 1.3.1 *Seja $S \subset X$ um subconjunto qualquer de um espaço vetorial X. Uma* **combinação linear** *de elementos de S é uma soma*

$$\lambda_1 x_1 + \ldots + \lambda_k x_k,$$

com $\lambda_1, \ldots, \lambda_k \in \mathbb{K}$ e $x_1, \ldots, x_k \in S$.

O conjunto S é **linearmente dependente** *se existe um número finito de elementos*

$$x_1, \ldots, x_k \in S$$

e escalares $\lambda_1, \ldots, \lambda_k \in \mathbb{K}$, não todos nulos, tais que

$$\lambda_1 x_1 + \ldots + \lambda_k x_k = 0.$$

Caso contrário, o conjunto S é **linearmente independente**.

O conjunto S **gera** *o espaço X se, para todo $x \in X$, existem (finitos) elementos $x_1, \ldots, x_j \in S$ tais que $x = \lambda_1 x_1 + \ldots + \lambda_j x_j$, para escalares $\lambda_1, \ldots, \lambda_j \in \mathbb{K}$. Uma* **base** *de X é um subconjunto S que é linearmente independente e gera X. Um espaço vetorial tem* **dimensão finita** *se tem uma base com um número finito de elementos.*

Lema 1.3.2 *Suponhamos que $S = \{x_1, \ldots, x_n\}$ gere o espaço vetorial X e que $\{y_1, \ldots, y_j\}$ seja linearmente independente em X. Então*

$$j \leq n.$$

Demonstração: Suponhamos $j > n$. Como S gera X, temos que

$$y_1 = \lambda_1 x_1 + \ldots + \lambda_n x_n,$$

sendo ao menos um dos escalares $\lambda_1, \ldots, \lambda_n$ diferente de zero (veja exercício 10). Podemos supor $\lambda_1 \neq 0$. Temos então que $\{x_2, \ldots, x_n, y_1\}$ gera X. De fato, se $x \in X$, existem escalares $\alpha_1, \ldots, \alpha_n$ tais que $x = \alpha_1 x_1 + \ldots + \alpha_n x_n$. Mas então

$$x = \alpha_1 \left[\frac{1}{\lambda_1} (y_1 - \lambda_2 x_2 - \ldots - \lambda_n x_n) \right] + \alpha_2 x_2 + \ldots + \alpha_n x_n,$$

mostrando o afirmado.

De maneira análoga, $y_2 = \beta_2 x_2 + \ldots + \beta_n x_n + \beta_1 y_1$, com ao menos um dos escalares $\beta_2, \ldots, \beta_n$ diferente de zero (veja o exercício 11). Supondo $\beta_2 \neq 0$, verificamos então que o conjunto $\{x_3, \ldots, x_n, y_1, y_2\}$ gera o espaço X. Repetindo sucessivamente esse procedimento, obtemos que

$$\{y_1, \ldots, y_n\}$$

gera o espaço X. Em particular,

$$y_{n+1} = \gamma_1 y_1 + \ldots + \gamma_n y_n.$$

Mas então

$$-\gamma_1 y_1 - \ldots - \gamma_n y_n + 1 y_{n+1} + 0 y_{n+2} + \ldots + 0 y_j = 0,$$

o que contradiz $\{y_1, \ldots, y_j\}$ ser um conjunto linearmente independente.

□

Lema 1.3.3 *Todo espaço vetorial gerado por um subconjunto finito $S = \{x_1, \ldots, x_n\}$ possui uma base.*

Demonstração: Se S é linearmente dependente, um de seus elementos pode ser escrito como combinação linear dos elementos restantes. Retirando esse elemento, o conjunto restante continua gerando X. Continuamos retirando elementos que são combinação linear dos elementos restantes até obter um conjunto linearmente independente que continua gerando X.

□

Um espaço vetorial de dimensão finita possui muitas bases.

Teorema 1.3.4 *Todas as bases de um espaço vetorial X de dimensão finita possuem o mesmo número de elementos.*

Demonstração: Se $S = \{x_1, \ldots, x_n\}$ e $S' = \{y_1, \ldots, y_j\}$ são bases de X, o lema 1.3.2 aplicado ao conjunto linearmente independente S' e ao conjunto gerador S mostra que $j \leq n$. Aplicando então ao conjunto linearmente independente S e ao conjunto gerador S', obtemos $n \leq j$.

□

Definição 1.3.5 *Se $S = \{x_1, \ldots, x_n\}$ é uma base do espaço vetorial X, dizemos que X tem dimensão n e escrevemos*

$$\dim X = n.$$

Se $X = \{0\}$, X tem dimensão finita igual a zero.

Teorema 1.3.6 *Todo subconjunto linearmente independente $S = \{y_1, \ldots, y_j\}$ de um espaço vetorial X de dimensão n pode ser completado para formar uma base de X.*

Demonstração: Se S não gera X, então existe um vetor $x_1 \in X$ que não é combinação linear dos elementos de S. O conjunto

$$\{y_1, \ldots, y_j, x_1\}$$

é linearmente independente. Repetimos esse procedimento um número finito de vezes, até obter uma base de X.

$\square$

O teorema 1.3.6 nos mostra como obter diferentes bases para um espaço vetorial de dimensão finita.

Observação 1.3.7 Uma base de um espaço vetorial é um conjunto ordenado. Assim, se $S = \{x_1, x_2, \ldots, x_n\}$ é uma base do espaço X, então $S' = \{x_2, \ldots, x_n, x_1\}$ é outra base de X. ◀

Definição 1.3.8 *Sejam X um espaço vetorial e $\mathcal{B} = \{x_1, \ldots, x_n\}$ uma base de X. Se $x \in X$, então existem escalares $\lambda_1, \ldots, \lambda_n \in \mathbb{K}$ tais que*

$$x = \lambda_1 x_1 + \ldots + \lambda_n x_n.$$

O vetor $(\lambda_1, \ldots, \lambda_n) \in \mathbb{K}^n$ é chamado **representação** *de x na base $\mathcal{B}$ e $\lambda_1, \ldots, \lambda_n$ as* **coordenadas** *de x na base $\mathcal{B}$. Denotamos também por $[x]_{\mathcal{B}}$ o vetor $(\lambda_1, \ldots, \lambda_n)$.*

Definição 1.3.9 *Seja $e_i \in \mathbb{K}^n$ o vetor cuja i-ésima coordenada é igual a 1, as outras sendo nulas. O conjunto $\mathcal{E} = \{e_1, \ldots, e_n\}$ é a* **base canônica** *do espaço $\mathbb{K}^n$.*

Teorema 1.3.10 *Seja X um espaço vetorial de dimensão finita. Então vale:*

(*i*) *todo subespaço Y de X possui dimensão finita;*

(*ii*) *todo subespaço Y possui um complemento $Z \subset X$, isto é, existe um subespaço Z de X tal que*

$$X = Y \oplus Z.$$

Demonstração: Se $Y = \{0\}$, então $\dim Y = 0$. Tome $0 \neq y_1 \in Y$. Se existir $y_2 \in Y$ linearmente independente com y_1, consideramos então o conjunto $\{y_1, y_2\}$. Se esse conjunto gera Y, temos uma base. Caso contrário, podemos acrescentar $y_3 \in Y$ linearmente independente com y_1 e y_2. Procedendo assim, obtemos sucessivamente conjuntos linearmente independentes, cada um contendo o anterior. De acordo com o lema 1.3.2, esse processo só pode continuar enquanto esses conjuntos tiverem dimensão menor do que a dimensão de X. Obtemos assim uma base $\{y_1, \ldots, y_j\}$ para Y.

Aplicando então o teorema 1.3.6, essa base pode ser completada até obtermos uma base $\{y_1, \ldots, y_j, x_1, \ldots, x_{n-j}\}$ para X. Defina Z como o espaço de todas as combinações lineares dos elementos $x_1, \ldots, x_{n-j}$. Claramente Z é um subespaço de X e $Z \cap Y = \{0\}$. Logo, pela proposição 1.2.4, temos $X = Y \oplus Z$.

$\square$

1.4 Espaço quociente

Definição 1.4.1 *Seja Y um subespaço de X. Se $x_1, x_2 \in X$, dizemos que x_1 é* **congruente** *a x_2 módulo Y, escrito*

$$x_1 \equiv x_2 \mod Y,$$

se $x_1 - x_2 \in Y$.

Podemos dividir o espaço X em diferentes classes de equivalência módulo Y (veja exercício 23). Denotaremos a classe contendo o elemento x por $[x]$.

Definição 1.4.2 *Se $[x]$ e $[z]$ são classes de equivalência módulo Y e $\lambda \in \mathbb{K}$, definimos*

$$[x] + [z] = [x + z], \quad \lambda[x] = [\lambda x].$$

Com essas operações, o conjunto de todas as classes de equivalência módulo Y torna-se um espaço vetorial, denotado

$$\frac{X}{Y} \quad \text{ou} \quad X/Y$$

e denominado **espaço quociente** *de X por Y.*

A classe de equivalência $[x]$ muitas vezes é representada por $x + Y$.

A rigor, precisamos mostrar que as operações em X/Y estão bem definidas, isto é, independem dos representantes de cada classe de equivalência. Portanto, suponhamos que $x_1 \in [x]$ e $z_1 \in [z]$. Então $x_1 = x + y_1$ e $z_1 = z + y_2$, com $y_1, y_2 \in Y$. Mas então $x_1 + z_1 = x + y_1 + z + y_2 = x + z + (y_1 + y_2)$ e, assim, $x_1 + z_1 \equiv x + z \mod Y$. Do mesmo modo, $\lambda x_1 = \lambda x + (\lambda y_1)$ e $\lambda x_1 \equiv \lambda x \mod Y$.

Exemplo 1.4.3 Seja $x \in \mathbb{K}^n$ e considere Y o subespaço de todos os vetores cujas duas primeiras coordenadas são nulas. Então dois vetores são congruentes módulo Y se, e somente se, suas duas primeiras coordenadas são iguais. Isto é,

$$(x_1, x_2, x_3, \ldots, x_n) \equiv (y_1, y_2, y_3, \ldots, y_n) \mod Y \quad \Leftrightarrow \quad x_1 = y_1 \text{ e } x_2 = y_2.$$

Cada classe de equivalência pode ser vista como um vetor com duas componentes, quais sejam, as duas coordenadas que eles possuem em comum. ◀

Teorema 1.4.4 *Seja Y um subespaço do espaço vetorial de dimensão finita X. Então*

$$\dim X = \dim Y + \dim \frac{X}{Y}.$$

Demonstração: Seja $\{y_1, \ldots, y_j\}$ uma base de Y. Podemos completá-la de modo que

$$\{y_1, \ldots, y_j, x_{j+1}, \ldots, x_n\}$$

seja uma base de X. Afirmamos que $\{x_{j+1}, \ldots, x_n\}$ é uma base de X/Y. De fato, se $v \in X/Y$, então $v = \lambda_1 y_1 + \ldots + \lambda_j y_j + \lambda_{j+1} x_{j+1} + \ldots + \lambda_n x_n$. Mas então $v = \lambda_{j+1} x_{j+1} + \ldots + \lambda_n x_n + y$, em que $y = \lambda_1 y_1 + \ldots + \lambda_j y_j \in Y$.

□

Temos então, imediatamente, o seguinte

Corolário 1.4.5 *Se Y é um subespaço de X e $\dim Y = \dim X$, então $Y = X$.*

1.5 Exercícios

1. Se $-x$ é o inverso aditivo de $x \in X$, mostre que $-x = (-1)x$.

2. Mostre que o elemento neutro aditivo de um espaço vetorial é único. Mostre que $0x = 0$ para todo $x \in X$ e $\lambda 0 = 0$ para todo $\lambda \in \mathbb{K}$, sendo $0 \in X$ o elemento neutro aditivo.

3. Mostre que $Y \subset X$ é um subespaço se, e somente se, $\lambda x + y \in Y$ para quaisquer $x, y \in Y$ e $\lambda \in \mathbb{K}$.

4. Se X é um espaço vetorial, mostre que os conjuntos X e $\{0\}$ (que consiste apenas do elemento neutro aditivo) são subespaços de X, chamados **subespaços triviais**.

5. Seja $X = \{(x_1, \ldots, x_n) : x_i \in \mathbb{K}\}$. Defina a soma $x + y$ da maneira usual e $\lambda x = 0$ para todo $\lambda \in \mathbb{K}$ e $x \in X$. Verifique quais propriedades da definição de espaço vetorial são satisfeitas.

6. Seja $V \subset \mathbb{K}^n$ o conjunto de todas as n-uplas da forma $(0, 0, x_3, \ldots, x_n)$. Mostre que V é um subespaço de $\mathbb{K}^n$.

7. Seja $U = \{(x, y) \in \mathbb{R}^2 : x > 0, y > 0\}$. Se $z_1 = (x_1, y_1)$ e $z_2 = (x_2, y_2)$ são elementos de U e $\lambda \in \mathbb{R}$, defina
$$z_1 + z_2 = (x_1x_2, y_1y_2), \quad \lambda z_1 = (x_1^\lambda, y_1^\lambda).$$
   (a) Mostre que U é um espaço vetorial;
   (b) mostre que, se $v_1 = (e, 1)$ e $v_2 = (1, e)$, então $\mathcal{B} = \{v_1, v_2\}$ é uma base de U (estamos denotando e a base dos logaritmos naturais).
   (c) Defina $T : U \to \mathbb{R}^2$ por $T(z) = [z]_\mathcal{B}$, em que $[z]_\mathcal{B}$ é a representação de z na base $\mathcal{B}$. Mostre que T é um isomorfismo.

8. Seja $S \subset X$ um subconjunto arbitrário do espaço vetorial X. Mostre que o conjunto de todas as combinações lineares dos elementos de S forma um subespaço de X, chamado **espaço gerado** por S e denotado $< S >$. Mostre que se $Y \subset X$ é um subespaço tal que $S \subset Y$, então $< S > \subset Y$. (Esse exercício generaliza o procedimento usado na demonstração do teorema 1.3.10).

9. Mostre que $U \cap V$ é um subespaço de X, se U e V são subespaços de X. O subespaço $U \cap V$ é chamado **interseção** dos subespaços U e V.

10. Se $S \subset X$ é linearmente independente, mostre que $0 \notin S$. Mostre que se um conjunto possui um subconjunto linearmente dependente, então esse conjunto é linearmente dependente.

11. Qual a razão, na demonstração do lema 1.3.2, de substituirmos sempre um dos elementos $x_j, \ldots, x_n$ do conjunto $\{x_j, \ldots, x_n, y_1, \ldots, y_{j-1}\}$ pelo elemento y_j? Porque não podemos substituir y_j por um dos elementos $y_1, \ldots, y_{j-1}$?

12. Seja $\mathcal{P}$ o espaço vetorial de todos os polinômios na variável x, com coeficientes em $\mathbb{K}$. Seja $S = \{1, x, x^2, \ldots, x^n, \ldots\}$. Mostre que S é uma base de $\mathcal{P}$.

13. Mostre que uma transformação linear $T : X \to Y$ é injetiva se, e somente se, $\ker T = \{0\}$, em que $\ker T := \{v \in X;\ Tv = 0\}$.

14. Mostre que $\mathbb{K}^n$ e $\mathcal{P}_n$ são isomorfos.

15. Seja $T : X \to Y$ um isomorfismo entre os espaços X e Y. Mostre que a **inversa** $T^{-1} : Y \to X$ é linear.

16. Mostre que todo espaço vetorial de dimensão n sobre o corpo $\mathbb{K}$ é isomorfo a $\mathbb{K}^n$. Esse isomorfismo é único? Conclua que quaisquer dois espaços de dimensão n sobre o mesmo corpo $\mathbb{K}$ são sempre isomorfos. Os espaços $\mathbb{R}^n$ e $\mathbb{C}^n$ são isomorfos?

17. Mostre que S é uma base de X se, e somente se, todo elemento $x \in X$ pode ser escrito de maneira única como combinação linear dos elementos de S.

18. Seja X um espaço vetorial de dimensão n. Se $S = \{y_1, \ldots, y_n\}$ é um conjunto linearmente independente, mostre que S é uma base de X.

19. Sejam X um espaço vetorial de dimensão n e $S = \{y_1, \ldots, y_n\}$ um conjunto que gera X. Mostre que S é uma base de X.

20. Seja X um espaço de dimensão n e $V_1 \oplus \cdots \oplus V_k$ uma soma direta de subespaços de X. Mostre que $\dim V_1 \oplus \cdots \oplus V_k = \dim V_1 + \ldots + \dim V_k \leq n$.

21. Sejam U, V subespaços de X. Mostre que $\dim U + V = \dim U + \dim V - \dim(U \cap V)$.

22. Denotaremos por $\mathbb{M}_{n\times n}$ o conjunto das matrizes $n \times n$. Defina $\mathcal{S} = \{A \in \mathbb{M}_{n\times n};\ A^T = A\}$, em que A^T denota a transposta da matriz A ($\mathcal{S}$ é o conjunto das matrizes **simétricas**); defina $\mathcal{A} = \{A \in \mathbb{M}_{n\times n};\ A^T = -A\}$ ($\mathcal{A}$ é o conjunto das matrizes anti-simétricas). Mostre que $\mathbb{M}_{n\times n} = \mathcal{S} \oplus \mathcal{A}$.

23. Seja $\sim$ uma relação de equivalência[1] num conjunto A. Dado $x \in A$, denote
$$cl(x) =: \{y \in A;\ y \sim x\}$$
a classe de equivalência do elemento x. Mostre que A pode ser escrito como uma união disjunta de suas classes de equivalência.

24. Mostre que a congruência módulo Y é uma relação de equivalência.

25. Seja $W \subset \mathbb{R}^3$ o **subespaço** (verifique!) formado por todas as soluções da equação linear homogênea $2x + 3y + 4z = 0$. Descreva as classes de equivalência de W em $\mathbb{R}^3$.

26. Seja Y um subespaço de X. Mostre que X é isomorfo a $Y \oplus X/Y$.

27. A **soma direta** de espaços vetoriais X_1, X_2 é o conjunto $X_1 \oplus X_2$ de todos os pares (x_1, x_2) com $x_1 \in X_1$ e $x_2 \in X_2$. Definindo adição e multiplicação por escalar coordenada a coordenada, mostre que $X_1 \oplus X_2$ é um espaço vetorial. Se X_1 e X_2 têm dimensão finita, então $\dim X_1 \oplus X_2 = \dim X_1 + \dim X_2$.

[1]Quer dizer, se $x, y, z \in A$, então: (i) $x \sim x$; (ii) se $x \sim y$, então $y \sim x$; (iii) se $x \sim y$ e $y \sim z$, então $x \sim z$.

Capítulo 2

Dualidade

O capítulo visa a apresentação de uma primeira versão do Teorema de Representação de Riesz e também do isomorfismo canônico entre o espaço X e o bidual X''. Ele pode ser suprimido numa primeira leitura ou a critério do instrutor.

2.1 Isomorfismos

Lema 2.1.1 *Sejam X, Y espaços vetoriais de dimensão finita sobre o corpo $\mathbb{K}$. Então, se $T : X \to Y$ é um isomorfismo, a imagem por T de toda base de X é uma base de Y. Em particular,* $\dim X = \dim Y$.

Demonstração: Seja $\{x_1, \ldots, x_n\}$ uma base de X. Afirmamos que $\{Tx_1, \ldots, Tx_n\}$ é uma base de Y. De fato, seja $y \in Y$ qualquer. Existe um único $x \in X$ tal que $Tx = y$. Mas $x = \lambda_1 x_1 + \ldots + \lambda_n x_n$ para escalares $\lambda_1, \ldots, \lambda_n \in \mathbb{K}$. A linearidade de T então garante que

$$y = T(x) = \lambda_1 Tx_1 + \ldots + \lambda_n Tx_n,$$

mostrando que $\{Tx_1, \ldots, Tx_n\}$ gera Y. Suponhamos agora que $\lambda_1 Tx_1 + \ldots + \lambda_n Tx_n = 0$ para certos escalares $\lambda_1, \ldots, \lambda_n$. Então $T(\lambda_1 x_1 + \ldots + \lambda_n x_n) = 0$. Como T é injetora, $\lambda_1 x_1 + \ldots + \lambda_n x_n = 0$. Como $\{x_1, \ldots, x_n\}$ é base, $\lambda_1 = \ldots = \lambda_n = 0$.

□

2.2 O espaço dual

Definição 2.2.1 *Se X é um espaço vetorial sobre $\mathbb{K}$, consideremos o conjunto*

$$\{\ell : X \to \mathbb{K} : \ell \text{ é linear}\}.$$

De maneira natural vemos que esse conjunto tem uma estrutura de espaço vetorial, se definirmos, para λ escalar e ℓ, m nesse conjunto,

$$(\ell + m)(x) = \ell(x) + m(x), \quad (\lambda\ell)(x) = \lambda\ell(x).$$

Com essas operações, denotamos $X' = \{\ell : X \to \mathbb{K} : \ell \text{ é linear}\}$ o **espaço dual** *de X. Os elementos de X' são chamados* **funcionais lineares**.

Exemplo 2.2.2 Seja $X = \{f : [0,1] \to \mathbb{R} : f$ é contínua$\}$. Defina $\ell(f) = \int_0^1 f(s)ds$ e, para $s_0 \in [0,1]$ fixo, $m(f) = f(s_0)$. Então $\ell, m \in X'$. ◀

Seja $\{x_1, \ldots, x_n\}$ uma base do espaço vetorial X. Para $x \in X$, existem escalares $\ell_1(x), \ldots, \ell_n(x)$ tais que

$$x = \ell_1(x)x_1 + \ldots + \ell_n(x)x_n.$$

Os escalares são justamente as coordenadas de x na base $\{x_1, \ldots, x_n\}$. (Quer dizer, se $x = \alpha_1 x_1 + \ldots + \alpha_n x_n$ e $y = \beta_1 x_1 + \ldots + \beta_n x_n$, estamos denotando $\ell_i(x) = \alpha_i$ e $\ell_i(y) = \beta_i$).

Teorema 2.2.3 *Seja $\mathcal{B} = \{x_1, \ldots, x_n\}$ uma base de X e*

$$x = \ell_1(x)x_1 + \ldots + \ell_n(x)x_n.$$

Então:

(*i*) *para todo $i = 1, \ldots, n$, $\ell_i : X \to \mathbb{K}$ é um funcional linear e $\ell_i(x_j) = \delta_{ij}$;*

(*ii*) *o conjunto $\{\ell_1, \ldots, \ell_n\}$ é uma base de X', chamada* **base dual** *da base $\mathcal{B}$;*

(*iii*) *se $m \in X'$, então*

$$m(x) = \ell_1(x)m(x_1) + \ldots + \ell_n(x)m(x_n).$$

(*iv*) *para todo $0 \neq x \in X$, existe $m \in X'$ tal que $m(x) \neq 0$.*

Demonstração: (*i*) Suponhamos que $x = \alpha_1 x_1 + \ldots + \alpha_n x_n$ e $y = \beta_1 x_1 + \ldots + \beta_n x_n$ (quer dizer, $\ell_i(x) = \alpha_i$ e $\ell_i(y) = \beta_i$). Então $x + \lambda y = (\alpha_1 + \lambda\beta_1)x_1 + \ldots + (\alpha_n + \lambda\beta_n)x_n$ e, portanto $\ell_i(x + \lambda y) = \alpha_i + \lambda\beta_i = \ell_i(x) + \lambda\ell_i(y)$.

(*ii*) Suponhamos que $\lambda_1\ell_1 + \ldots + \lambda_n\ell_n = 0 \in X'$. Avaliando esse funcional sucessivamente nos vetores $x_1, \ldots, x_n$ concluímos que $\lambda_1 = \ldots = \lambda_n = 0$. Seja agora $m \in X'$. Então

$$m(x) = m(\alpha_1 x_1 + \ldots + \alpha_n x_n) = \alpha_1 m(x_1) + \ldots + \alpha_n m(x_n) = \ell_1(x)m(x_1) + \ldots + \ell_n(x)m(x_n),$$

provando não apenas que $\ell_1, \ldots, \ell_n$ gera X', mas também a afirmação (*iii*).

(*iv*) Se $0 \neq x$, então alguma coordenada $\ell_i(x)$ na expressão $x = \ell_1(x)x_1 + \ldots + \ell_n(x)x_n$ não é nula. Considere $m = \ell_i$.

□

Observação 2.2.4 A parte (*iii*) do teorema 2.2.3 é uma versão do Teorema de Representação de Riesz; veja o teorema 6.3.5. ◀

Uma vez que X' é um espaço vetorial de dimensão n, temos que esse espaço tem o seu dual, que denotaremos X'' e chamaremos o **bidual** de X. O teorema anterior garante então que $\dim X'' = n$, pois já vimos que $\dim X' = n$.

Note que X'' é, por definição, o espaço vetorial de aplicações lineares

$$X'' = \{L : X' \to \mathbb{K} : L \text{ é linear}\}.$$

Quer dizer, L é uma transformação linear que associa, a cada funcional linear $\ell : X \to \mathbb{K}$, o número $L(\ell) \in \mathbb{K}$. Os elementos de X'' são, aparentemente, complicados. Mostraremos que as aplicações lineares em X'' estão canonicamente associadas aos vetores do espaço X. Quer dizer, existe um isomorfismo entre X e X'' que independe da utilização de qualquer base nesses espaços vetoriais. (A existência de um isomorfismo entre esses espaços é trivial, já que eles têm a mesma dimensão; veja o exercício 16 do capítulo 1).

Lema 2.2.5 *Para cada $x \in X$ fixo, considere a aplicação $L_x : X' \to \mathbb{K}$ definida por*

$$L_x(\ell) = \ell(x).$$

Quer dizer, L_x associa a cada funcional linear $\ell \in X'$ o valor que ℓ assume no ponto x. Então $L_x \in X''$.

Demonstração: Suponhamos que $\ell, m \in X'$. Então, se $\alpha \in \mathbb{K}$,

$$L_x(\ell + \alpha m) = (\ell + \alpha m)(x) = \ell(x) + \alpha m(x) = L_x(\ell) + \alpha L_x(m).$$

(Compare essa demonstração com o exemplo 2.2.2).

$\square$

Teorema 2.2.6 *Todo elemento do espaço X'' é da forma L_x, para algum $x \in X$.*

Demonstração: Apesar de ser constituída de etapas bastante simples, a idéia da demonstração é relativamente elaborada. Definimos $\Gamma = \{L_x : x \in X\}$. Quer dizer, os elementos de Γ são as aplicações lineares definidas no lema anterior. Vamos mostrar, em primeiro lugar, que Γ é um subespaço de X''. Depois, mostraremos que X é isomorfo a Γ. Assim, $\dim \Gamma = n = \dim X''$. Isso quer dizer que $\Gamma = X''$.
1a. parte: Γ é um subespaço de X''.

Sejam $L_x, L_y \in \Gamma$ e $\lambda \in \mathbb{K}$. Consideremos $L_x + \lambda L_y$. Queremos mostrar que essa aplicação linear é um elemento de Γ, isto é, $L_x + \lambda L_y = L_z$ para algum $z \in X$. Temos, para $\ell \in X'$,

$$(L_x + \lambda L_y)(\ell) = L_x(\ell) + \lambda L_y(\ell) = \ell(x) + \lambda \ell(y) = \ell(x + \lambda y) = L_{x+\lambda y}(\ell).$$

2a. parte: X é isomorfo a Γ. Definimos

$$\begin{array}{rccc} T: & X & \to & \Gamma \\ & x & \mapsto & L_x. \end{array}$$

Vamos mostrar que T é um isomorfismo entre X e Γ. Temos que

$$T(x + \lambda y) = L_{x+\lambda y} = L_x + \lambda L_y = T(x) + \lambda T(y),$$

de acordo com o que mostramos na primeira parte. A aplicação T é sobrejetiva por definição. A injetividade também é clara: se $T(x) = T(y)$, então $L_x = L_y$ e, portanto, $L_x(\ell) = L_y(\ell)$ para todo $\ell \in X'$. Mas então $\ell(x) = \ell(y)$ e $\ell(x - y) = 0$ para todo $\ell \in X'$. Mas isto implica que $x - y = 0$, de acordo com o teorema 2.2.3, (iv). Isto mostra a injetividade e completa a demonstração.

$\square$

Concluímos esse capítulo com o seguinte resultado, surpreendente à primeira vista:

Teorema 2.2.7 *Sejam $t_1, \ldots, t_n$ pontos distintos do intervalo I. Então existem constantes $m_1, \ldots, m_n$ tais que*

$$\int_I p(t)dt = m_1 p(t_1) + \ldots + m_n p(t_n)$$

para todo polinômio p de grau menor do que n.

Demonstração: O espaço $\mathcal{P}_n$ de todos os polinômios $p(t) = a_0 + a_1 t + \ldots + a_{n-1}t^{n-1}$ de grau menor do que n é isomorfo a $\mathbb{K}^n$ e, portanto, tem dimensão n.

Definimos $\ell_j(p) = p(t_j)$. Então $\ell_j \in \mathcal{P}_n'$. Afirmamos que $\{\ell_1, \ldots, \ell_n\}$ é linearmente independente. De fato, suponhamos que

$$\lambda_1 \ell_1 + \ldots + \lambda_n \ell_n = 0 \in \mathcal{P}_n'.$$

Isso implica que

$$\lambda_1 p(t_1) + \ldots + \lambda_n p(t_n) = 0, \quad \forall\, p \in \mathcal{P}_n. \tag{2.1}$$

Considere os polinômios

$$q_1(t) = (t - t_2) \cdots (t - t_n),\ q_2(t) = (t - t_1)(t - t_3) \cdots (t - t_n), \ldots, q_n(t) = (t - t_1) \ldots (t - t_{n-1}).$$

Cada polinômio q_i possui exatamente $n - 1$ raízes nos pontos t_j, com $j \neq i$. Substituindo sucessivamente os polinômios q_i na relação (2.1), obtemos $\lambda_i q(t_i) = 0$, o que implica $\lambda_i = 0$. Isso mostra que $\{\ell_1, \ldots, \ell_n\}$ é linearmente independente em $\mathcal{P}_n'$ e, portanto, uma base desse espaço, que tem dimensão n.

Assim, todo funcional linear $\ell : \mathcal{P}_n \to \mathbb{R}$ é uma combinação linear dos funcionais $\ell_1, \ldots, \ell_n$ e, portanto,

$$\ell = m_1 \ell_1 + \ldots + m_n \ell_n$$

para escalares $m_1, \ldots, m_n \in \mathbb{K}$. O resultado decorre ao considerarmos o funcional linear

$$p \mapsto \int_I p(t)dt.$$

□

2.3 Exercícios

1. Considere a base $\mathcal{B} := \{v_1, v_2\}$ do $\mathbb{R}^2$, em que $v_1 = (2, 1)$ e $v_2 = (3, 1)$. Acha a base dual de $\mathcal{B}$.

2. Seja $\mathcal{P}_n$ o espaço de todos os polinômios (com coeficientes em $\mathbb{R}$) de grau menor do que n. Mostre que as seguintes aplicações pertencem ao dual de $\mathcal{P}_n$: (a) $\pi_i(p(t)) = a_i$ para todo $i = 0, 1, \ldots, n - 1$, se $p(t) \in \mathcal{P}_n$ é dado por $p(t) = a_0 + a_1 t + \ldots + a_{n-1}t^{n-1}$; (b) $J(p(t)) = \int_0^1 p(t)dt$, para todo $p(t) \in \mathcal{P}_n(t)$.

3. Considere o espaço $\mathcal{P}_2$, como acima. Sejam $\ell_1 : \mathcal{P}_2 \to \mathbb{R}$ e $\ell_2 : \mathcal{P}_2 \to \mathbb{R}$ dadas por $\ell_1(p(t)) = \int_0^1 p(t)dt$ e $\ell_2(p(t)) = \int_0^2 p(t)dt$. Mostre que $\mathcal{B}' = \{\ell_1, \ell_2\}$ é uma base de $\mathcal{P}_2'$. Ache a base $\{v_1, v_2\}$ de $\mathcal{P}_2$ da qual $\mathcal{B}'$ é dual.

4. Sejam X um espaço vetorial arbitrário e $f : X \to \mathbb{K}$ um funcional linear não-nulo.

   (a) Mostre que $\ker f$ tem **codimensão 1**, isto é, existe $w \in X$ tal que
   $$X = \ker f \oplus < w > .$$
   (denotamos $< w >$ o espaço gerado por $w \in X$).

   (b) Se $g : X \to \mathbb{K}$ é outro funcional linear, então g é um múltiplo escalar de f se, e somente se, o núcleo de g contiver o núcleo de f.

   (c) Sejam $\varphi, f_1, \ldots, f_r$ funcionais lineares no espaço X. Mostre que φ é combinação linear de $f_1, \ldots, f_r$ se, e somente se, $\ker f_1 \cap \cdots \cap \ker f_r \subset \ker \varphi$.

5. Sejam X um espaço vetorial e $S \subset X$ um subconjunto arbitrário. O **anulador** de S é o conjunto $S^\perp = \{f \in X' : f(s) = 0 \ \forall \ s \in S\}$. Mostre que $S^\perp$ é um subespaço de X'.

6. Seja $Y \subset X$ um subespaço do espaço vetorial de dimensão finita X. Mostre que $\dim X = \dim Y + \dim Y^\perp$. Identificando X e X'' (de acordo com o teorema 2.2.6), mostre que $Y^{\perp\perp} = Y$.

7. Seja $S = \{(2, -2, 3, 4, -1), (-1, 1, 2, 5, 2), (0, 0, -1, -2, 3), (1, -1, 2, 3, 0)\} \subset \mathbb{R}^5$. Obtenha o anulador de $< S >$.

8. Sejam A, B matrizes $n \times n$. Mostre que a igualdade $AB - BA = I$ nunca é satisfeita.

9. Seja $W \subset X$ um subespaço e $f : W \to \mathbb{K}$ linear. Mostre que existe um funcional linear $\varphi : X \to \mathbb{K}$ que estende f, isto é, $\varphi(w) = f(w)$ para todo $w \in W$.

Capítulo 3

Aplicações Lineares

3.1 Aplicações lineares e matrizes I

Sejam X e Y espaços vetoriais sobre o mesmo corpo $\mathbb{K}$. Como sabemos, uma **aplicação linear** (ou transformação linear) é uma aplicação $T : X \to Y$ tal que

$$T(x + \lambda y) = T(x) + \lambda T(y), \quad \forall\, x, y \in X \text{ e } \lambda \in \mathbb{K}.$$

Exemplo 3.1.1 Um isomorfismo é sempre uma transformação linear. Se $X = Y = \mathbb{R}^2$, definindo T como sendo uma rotação de um ângulo θ em torno da origem, vemos que T é linear (verifique!). Se $\mathcal{P}$ é o espaço vetorial de polinômios, $T : \mathcal{P} \to \mathcal{P}$ definida por $T(p) = p'$ (derivação) é uma transformação linear, bem como $S(p) = \int p$ (integração). ◀

Exemplo 3.1.2 Sejam $X = \mathbb{R}^n$ e $Y = \mathbb{R}^m$ e $a_{ij} \in \mathbb{R}$, para $j = 1, \ldots, n$ e $i = 1, \ldots, m$. Para $x \in X$, definimos $y = Tx$ por

$$y_i = \sum_{j=1}^{n} a_{ij} x_j, \quad i = 1, \ldots, m. \tag{3.1}$$

(Estamos denotando $x = (x_1, \ldots, x_n)$ e $y = (y_1, \ldots, y_m)$, sendo $y_i = (Tx)_i$ a i-ésima coordenada de y). Afirmamos que T é linear. De fato, se $w = (w_1, \ldots, w_n) \in \mathbb{R}^n$ e $\lambda \in \mathbb{R}$, temos

$$(T(x + \lambda w))_i = \sum_{j=1}^{n} a_{ij}(x_j + \lambda w_j) = \sum_{j=1}^{n} a_{ij} x_j + \lambda \sum_{j=1}^{n} a_{ij} w_j = (Tx)_i + \lambda (Tw)_i.$$

(Escolha $i \in \{1, \ldots, m\}$ e escreva explicitamente a soma que está sendo efetuada). ◀

Teorema 3.1.3 *Toda aplicação linear $T : \mathbb{R}^n \to \mathbb{R}^m$ é da forma* (3.1).

Demonstração: Considere a base canônica $\{e_1, \ldots, e_n\}$ do $\mathbb{R}^n$. Temos então que $x = x_1 e_1 + \ldots + x_n e_n = \sum_{j=1}^{n} x_j e_j$. Como T é linear,

$$y = Tx = T\left(\sum_{j=1}^{n} x_j e_j\right) = \sum_{j=1}^{n} x_j T(e_j).$$

Denote a i-ésima coordenada do vetor $T(e_j)$ por a_{ij}, isto é, $a_{ij} = (T(e_j))_i$. Assim, a i-ésima coordenada de y é

$$y_i = \sum_{j=1}^{n} x_j a_{ij},$$

como queríamos provar.

□

É conveniente representar os coeficientes (a_{ij}) da expressão (3.1) como um arranjo retangular:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix};$$

denominamos tal arranjo **matriz** $m \times n$, m sendo o número de linhas e n o número de colunas. O elemento t_{ij} é a **entrada** correspondente à linha i e coluna j.

Definição 3.1.4 *Sejam T, S aplicações lineares de X para Y. Definimos*

$$(T + S)(x) = Tx + Sx, \quad (\lambda T)(x) = \lambda Tx.$$

Com essas operações, o conjunto de todas as aplicações lineares de X para Y é um espaço vetorial, denotado $\mathcal{L}(X, Y)$.

(Compare a definição acima com a definição do espaço dual).

O exemplo 3.1.2 e o teorema 3.1.3 mostram que existe uma correspondência bijetiva entre o conjunto de matrizes $m \times n$ e $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. Denotamos o elemento a_{ij} da matriz A, chamada **matriz que representa** T (com relação às bases canônicas do $\mathbb{R}^n$ e $\mathbb{R}^m$) por

$$T_{ij} = a_{ij}.$$

Lema 3.1.5 *Sejam $S, T : \mathbb{R}^n \to \mathbb{R}^m$. Então $(S + T)_{ij} = S_{ij} + T_{ij}$ e $(\lambda T)_{ij} = \lambda T_{ij}$.*

Em outras palavras, estão assim definidas a soma de duas matrizes $m \times n$ (como a matriz obtida ao se somar as entradas correspondentes de cada matriz) e a multiplicação de uma matriz por um escalar (como a matriz obtida ao se multiplicar cada entrada da matriz pelo escalar). As operações no espaço $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ correspondem às operações no conjunto das matrizes $m \times n$, fazendo desse conjunto, denotado $\mathbb{M}_{m\times n}$, um espaço vetorial.

Demonstração: Utilizando a notação do teorema 3.1.3, temos, por definição, que a_{ij} e b_{ij} são as i-ésimas coordenadas dos vetores $T(e_j)$ e $S(e_j)$. Assim, se somamos as i-ésimas coordenadas desses vetores, obtemos $b_{ij} + a_{ij}$. Por outro lado, $S(e_j) + T(e_j) = (S + T)(e_j)$, de modo que a i-ésima componente do vetor $(S + T)(e_j)$ é $b_{ij} + a_{ij}$.

Do mesmo modo, a i-ésima componente do vetor $(\lambda T)(e_j)$ é λ multiplicado pela i-ésima componente do vetor $T(e_j)$.

□

3.2 Composta de aplicações lineares e multiplicação de matrizes

Sejam X, Y e Z espaços vetoriais sobre o mesmo corpo $\mathbb{K}$, e $T : X \to Y$ e $S : Y \to Z$ aplicações lineares. Denotamos $S \circ T : X \to Z$ a aplicação composta de T com S. Quer dizer,

$$(S \circ T)x = S(Tx).$$

É fácil verificar que $S \circ T \in \mathcal{L}(X, Z)$. Além disso, se $R : Z \to W$ é linear, temos que $R \circ (S \circ T) = (R \circ S) \circ T$ (quer dizer, a composição de aplicações é uma operação associativa; isso independe da linearidade de R, S e T).

Mais do que isso, temos

$$(P + S) \circ T = P \circ T + S \circ T \quad \forall\ P \in \mathcal{L}(Y, Z)$$

e

$$S \circ (T + Q) = S \circ T + S \circ Q, \quad \forall\ Q \in \mathcal{L}(X, Y).$$

(também essas propriedades independem da linearidade).

É usual denotar, no caso de aplicações lineares, $S \circ T$ por ST e chamá-lo **produto** das aplicações lineares S e T. Note que, em geral, $ST \neq TS$ (na verdade, os dois lados nem precisam estar simultaneamente definidos; mesmo estando, não há razão para eles serem iguais).

Através do Lema 3.1.5 foram interpretadas as operações no espaço vetorial $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ em termos de operações entre matrizes, introduzindo assim operações em $\mathbb{M}_{m\times n}$ que fazem desse conjunto um espaço vetorial, isomorfo ao espaço $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ (verifique que temos realmente um isomorfismo!). A composição de aplicações lineares (quando possível) também pode ser interpretada como operação entre matrizes. Veremos que elas correspondem à multiplicação dessas, o que justifica a denominação de produto para a composição de aplicações lineares e a notação ST ao invés de $S \circ T$.

O nosso ponto de partida, para isso, consiste da expressão (3.1). Considerando o vetor $x = e_j$, vemos que o lado direito de (3.1) produz a j-ésima coluna da matriz (a_{ij}). Mas Te_j é justamente um vetor do $\mathbb{R}^m$, cuja i-ésima coordenada é a_{ij}. Assim, é natural interpretar os vetores em $\mathbb{R}^m$ como **colunas**. Para sermos consistentes com esse fato, interpretaremos tanto os vetores no $\mathbb{R}^n$ como os vetores no $\mathbb{R}^m$ como "vetores coluna".

Uma matriz A pode ser concebida de duas maneiras diferentes: como uma linha de vetores coluna ou como uma coluna de vetores linha:

$$A = (c_1\ c_2\ \ldots\ c_n) = \begin{pmatrix} \ell_1 \\ \vdots \\ \ell_m \end{pmatrix}, \quad \text{em que} \quad c_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{pmatrix} \quad \text{e} \quad \ell_i = (a_{i1}\ a_{i2}\ \cdots\ a_{in}). \tag{3.2}$$

Enfatizamos a equação obtida acima:

$$Te_j = c_j. \tag{3.3}$$

Utilizaremos as diversas concepções de uma matriz - arranjo de números ou de vetores linha ou vetores coluna - para podermos interpretar a composição de aplicações lineares e introduzirmos a multiplicação de matrizes.

Para isso, começamos por considerar um caso simples: aquele em que a matriz é composta por uma única linha. De acordo com o lema 3.1.5, uma matriz linha $(c_1 \ \dots \ c_n)$ corresponde a uma aplicação linear $\ell : \mathbb{R}^n \to \mathbb{R}$. Temos, assim, uma interpretação para os elementos do espaço dual do $\mathbb{R}^n$: eles são as matrizes-linha, isto é, as matrizes formadas por uma única linha e n colunas!

Calculando $\ell(x) = \ell x$ (o vetor x sendo interpretado como um vetor coluna), obtemos, de acordo com (3.1),

$$\ell x = (c_1 \ \dots \ c_n) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = c_1 x_1 + c_2 x_2 + \ldots + c_n x_n. \tag{3.4}$$

Essa fórmula, em particular, define o produto de uma matriz linha por uma matriz coluna!

A fórmula de multiplicação de uma matriz $m \times n$ por uma matriz coluna $n \times 1$ decorre também imediatamente de (3.1): se $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ é representada pela matriz (a_{ij}), então $y = Tx$ tem coordenadas

$$y_i = \sum_{j=1}^{n} a_{ij} x_j, \quad i = 1, \ldots, m. \tag{3.5}$$

Uma vez que já convencionamos que os nossos vetores são representados por colunas e

$$Tx = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

vemos que

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = Tx = \begin{pmatrix} \ell_1 \\ \ell_2 \\ \vdots \\ \ell_m \end{pmatrix} x = \begin{pmatrix} \ell_1 x \\ \ell_2 x \\ \vdots \\ \ell_m x \end{pmatrix}, \tag{3.6}$$

o que vem da comparação de (3.5) com (3.4).

Agora é fácil obter a fórmula de multiplicação de uma matriz $p \times m$ por uma matriz $m \times n$: uma matriz $p \times m$ corresponde a uma aplicação linear $S \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^p)$ e uma matriz $m \times n$ a uma aplicação linear $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. A composição $ST \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ está bem definida e produz uma matriz $p \times n$. Vamos caracterizar essa matriz. Pela equação (3.3), Te_j é igual a c_j, a j-ésima coluna de T. Do mesmo modo $(ST)e_j$ corresponde à j-ésima coluna da matriz que representa ST. Aplicando a fórmula (3.6) para $x = c_j = Te_j$, temos então

$$(ST)e_j = S(Te_j) = Sc_j = \begin{pmatrix} \ell_1 c_j \\ \vdots \\ \ell_p c_j \end{pmatrix},$$

em que ℓ_k é a k-ésima linha de S. Mostramos assim a regra: se S é uma matriz $p \times m$ e T uma matriz $m \times n$, então o produto ST é uma matriz $p \times n$, cuja entrada kj é o produto da k-ésima linha de S pela j-ésima coluna de T:

$$(ST)_{kj} = \ell_k c_j,$$

em que

$$S = \begin{pmatrix} \ell_1 \\ \vdots \\ \ell_k \end{pmatrix} \quad \text{e} \quad T = (c_1 \ \cdots \ c_n).$$

Note que, uma vez que o produto de transformações lineares é associativo, a multiplicação de matrizes é associativa. Outras propriedades básicas da multiplicação de matrizes decorrem, do mesmo modo, das propriedades análogas da composição de aplicações lineares.

3.3 O teorema do núcleo e da imagem

Definição 3.3.1 *Seja $T : X \to Y$ uma aplicação linear. Definimos a* **imagem** *de T, denotada $\mathcal{I}m\, T$, por*

$$\mathcal{I}m\, T := \{y \in Y;\ y = Tx\}.$$

Definimos o **núcleo** *de T, denotado* $\ker T$, *por*

$$\ker T := \{x \in X;\ Tx = 0\}.$$

O núcleo e a imagem de T são subespaços vetoriais de X e Y, respectivamente. De fato, se $x_1, x_2 \in \ker T$ e $\lambda \in \mathbb{K}$, então $T(x_1 + \lambda x_2) = T(x_1) + \lambda T(x_2) = 0 + \lambda 0 = 0$, provando que $x_1 + \lambda x_2 \in \ker T$. Se $y_1, y_2 \in \mathcal{I}m\, T$, então existem $x_1, x_2 \in X$ tais que $y_1 = T(x_1)$ e $y_2 = T(x_2)$. Logo, se $\lambda \in \mathbb{K}$, $y_1 + \lambda y_2 = T(x_1) + \lambda T(x_2) = T(x_1 + \lambda x_2)$, o que mostra que $y_1 + \lambda y_2 \in \mathcal{I}m\, T$.

Temos então um dos resultados mais importantes da Álgebra Linear:

Teorema 3.3.2 (do núcleo e da imagem) *Sejam X e Y espaços vetoriais de dimensão finita e $T \in \mathcal{L}(X, Y)$. Então*

$$\dim X = \dim \ker T + \dim \mathcal{I}m\, T.$$

Apresentaremos duas demonstrações distintas desse teorema. A primeira usa a linguagem de espaço quociente e é bastante sintética. A segunda é bastante construtiva.

Para motivar a primeira demonstração, cujo fundamento perpassa o estudo de todas as estruturas algébricas, apresentamos o

Exemplo 3.3.3 Seja A uma matriz $m \times n$ e considere o sistema linear não homogêneo $Ax = b$. Suponhamos que x_p seja uma solução desse sistema. Claramente, $x_p + z$ também é solução desse sistema para qualquer $z \in \ker A$. Mas essas são as únicas soluções. De fato, se x é outra solução, temos que $A(x - x_p) = 0$, de modo que $x - x_p = z \in \ker A$.

A igualdade $x = x_p + z$, com $z \in \ker A$, significa que $x \equiv x_p \mod \ker A$. Portanto, no espaço quociente $\mathbb{R}^n / \ker A$ a equação $Ax = b$ terá solução única $[x_p]$! ◀

1a. Demonstração: Essa prova pode ser sintetizada pelo seguinte diagrama:

$$\begin{array}{ccc} X & \overset{T}{\longrightarrow} & \mathcal{I}m\,T \subset Y \\ \downarrow & \nearrow & \\ \dfrac{X}{\ker T} & T_q & \end{array}$$

Vamos definir um isomorfismo $T_q : \frac{X}{\ker T} \to \mathcal{I}m\,T$. Como espaços isomorfos de dimensão finita têm a mesma dimensão, deduzimos que

$$\dim\left(\frac{X}{\ker T}\right) = \dim \mathcal{I}m\,T.$$

Mas, como já vimos, $\dim X/\ker T = \dim X - \dim \ker T$, de onde segue o teorema.

Definimos, para $[x] \in X/\ker T$, $T_q([x]) = Tx$. Temos:

1. T está bem definida: $x \equiv y \mod \ker T$ quer dizer que $T(x-y) = 0$, ou seja, $T(x) = T(y)$.

2. T_q é linear: $T_q([x] + \lambda[y]) = T_q([x + \lambda y]) = T(x + \lambda y) = Tx + \lambda Ty = T_q([x]) + \lambda T_q([y])$.

3. T_q é injetiva: se $T_q([x]) = T_q([y])$, então $Tx = Ty$ e $T(x-y) = 0$, donde $x \equiv y \mod \ker T$.

4. T_q é sobrejetiva, por definição.

Logo T_q é um isomorfismo e o resultado está provado.

□

A demonstração acima é a própria essência da utilidade do espaço quociente. Ela mostra que, mesmo se T não tiver inversa, podemos construir, de maneira natural, um isomorfismo à partir de T, no caso, a aplicação T_q.

2a. Demonstração: Como $\mathcal{I}m\,T \subset Y$ é um espaço vetorial de dimensão finita, existe uma base $\{y_1, \ldots, y_j\}$ para $\mathcal{I}m\,T$. Para cada elemento y_i existe $x_i \in X$ tal que $Tx_i = y_i$, com $1 \leq i \leq j$.

Afirmamos que o conjunto $\{x_1, \ldots, x_j\}$ assim obtido é linearmente independente. De fato, suponhamos que $\lambda_1 x_1 + \ldots + \lambda_j x_j = 0$. Então

$$0 = T(\lambda_1 x_1 + \ldots + \lambda_j x_j) = \lambda_1 T(x_1) + \ldots + \lambda_j T(x_j) = \lambda_1 y_1 + \ldots + \lambda_j y_j.$$

Como $y_1, \ldots, y_j$ são linearmente independentes, $\lambda_i = 0$ para $1 \leq i \leq j$, como queríamos.

Consideremos agora uma base $\{w_1, \ldots, w_k\}$ do núcleo de T. Afirmamos que

$$\{x_1, \ldots, x_j, w_1, \ldots, w_k\}$$

é uma base de X.

Dado $x \in X$, como $Tx \in \mathcal{I}m\,T$, $Tx = \lambda_1 y_1 + \ldots + \lambda_j y_j$, quer dizer, $Tx = T(\lambda_1 x_1 + \ldots + \lambda_j x_j)$ e portanto $T(x - \lambda_1 x_1 - \ldots - \lambda_j x_j) = 0$. Assim, $x - \lambda_1 x_1 - \ldots - \lambda_j x_j \in \ker T$, donde

$$x - \lambda_1 x_1 - \ldots - \lambda_j x_j = \alpha_1 w_1 + \ldots + \alpha_k w_k.$$

Isso mostra que $x = \lambda_1 x_1 + \ldots + \lambda_j x_j + \alpha_1 w_1 + \ldots + \alpha_k w_k$, e que $\{x_1, \ldots, x_j, w_1, \ldots, w_k\}$ gera X.

Suponhamos agora que $\lambda_1 x_1 + \ldots + \lambda_j x_j + \alpha_1 w_1 + \ldots + \alpha_k w_k = 0$. Aplicando T nessa igualdade temos $\lambda_1 y_1 + \ldots + \lambda_j y_j = 0$, o que nos permite concluir que $\lambda_i = 0$ para $i = 1, \ldots, j$. Mas então $\alpha_1 w_1 + \ldots + \alpha_k w_k = 0$. Como $\{w_1, \ldots, w_k\}$ é linearmente independente, temos $\alpha_i = 0$ para $i = 1, \ldots, k$, o que mostra que todos os escalares são nulos e completa a demonstração.

$\square$

Se você comparar essas duas demonstrações, você perceberá que a essência da segunda é o procedimento aplicado na primeira: mostrou-se que existe um isomorfismo entre $\mathcal{I}m\, T$, espaço cuja base é $\{y_1, \ldots, y_j\} = \{Tx_1, \ldots, Tx_j\}$, e o espaço gerado por $\{x_1, \ldots, x_j\}$. Esse último espaço é justamente $X/\ker T$!

Mostraremos agora algumas conseqüências do Teorema do núcleo e da imagem. As demonstrações seguem imediatamente da fórmula

$$\dim X = \dim \mathcal{I}m\, T + \dim \ker T.$$

Corolário 3.3.4 *Suponhamos que* $\dim Y < \dim X$. *Então existe* $x \neq 0$ *tal que* $Tx = 0$.

Demonstração: Note que, em particular, $\dim \mathcal{I}m\, T < \dim X$.

$\square$

O corolário 3.3.4 é muitas vezes formulado em termos de sistemas lineares:

Corolário 3.3.5 *Seja* $T : \mathbb{R}^n \to \mathbb{R}^m$ *linear, com* $m < n$. *Então o sistema linear homogêneo* $Tx = 0$ (*em que* T *está sendo identificada com a matriz que a representa*) *possui solução não trivial, isto é, existe* $x \neq 0$ *tal que* $Tx = 0$.

Corolário 3.3.6 *Se* $\dim X = \dim Y$, *então* T *é injetiva se, e somente se,* T *é sobrejetiva.*

Demonstração: Se T é injetiva, $T(x) = 0$ implica $x = 0$. Logo, $\dim \ker T = 0$. Assim, $\dim \mathcal{I}m\, T = \dim X = \dim Y$ e, portanto, $\mathcal{I}m\, T = Y$. Reciprocamente, se T é sobrejetiva, $\mathcal{I}m\, T = Y$ e, portanto, $\dim \ker T = 0$.

$\square$

Em particular o corolário 3.3.6 garante, quando $\dim X = \dim Y$, que T é injetiva se, e somente se, $\ker T = \{0\}$. Esse resultado é válido, na verdade, para quaisquer espaços vetoriais X e Y. De fato[1], se T é injetiva, claramente $\ker T = \{0\}$; se existisse $x_1 \neq x_2$ tal que $T(x_1) = T(x_2)$, então $T(x_1 - x_2) = 0$, com $x_1 - x_2 \neq 0$.

A formulação do corolário 3.3.6 em termos de sistemas lineares é a seguinte:

Corolário 3.3.7 *Seja* $T : \mathbb{R}^n \to \mathbb{R}^n$ *linear. Então o sistema não homogêneo* $Tx = y$ *tem solução única para todo* $y \in Y$ *se, e somente se, o sistema homogêneo* $Tx = 0$ *tem solução única.*

Finalmente, enunciamos o resultado apresentado no exemplo 3.3.3, que não passa de uma caracterização do isomorfismo dado na primeira demonstração do teorema do núcleo e da imagem:

[1] Veja exercício 13 do capítulo 1.

Proposição 3.3.8 *Seja $y \in \mathbb{R}^m$ um elemento da imagem de $T : \mathbb{R}^n \to \mathbb{R}^m$. Então existe um único elemento $x_p \in \mathbb{R}^n$ tal que toda solução de $Tx = y$ é congruente a x_p módulo* $\ker T$*, isto é, se $Tx = y$, então $x = x_p + z$, para algum $z \in \ker T$.*

3.4 O espaço linha e o espaço coluna de uma matriz

Como vimos, dada uma matriz $A = (a_{ij})$, podemos vê-la através de suas linhas ou colunas:

$$A = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ldots & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix} = (c_1 \ \ldots \ c_n) = \begin{pmatrix} \ell_1 \\ \vdots \\ \ell_m \end{pmatrix}. \tag{3.7}$$

Os vetores colunas $c_1, \ldots, c_n$ são vetores do $\mathbb{R}^m$. Se $\mathcal{C} = \{c_1, \ldots, c_n\}$, chamamos de espaço-coluna o espaço gerado por $\mathcal{C}$, isto é, $< \mathcal{C} > \subset \mathbb{R}^m$.

Por outro lado, podemos interpretar as linhas de A ou como elementos do dual $(\mathbb{R}^n)'$ ou como elementos do próprio espaço $\mathbb{R}^n$. Se escrevemos $\mathcal{L} = \{\ell_1, \ldots, \ell_m\} \subset \mathbb{R}^n$, chamamos de espaço-linha o espaço gerado por $\mathcal{L}$, isto é, $< \mathcal{L} > \subset \mathbb{R}^n$.

Começamos interpretando o espaço-coluna de uma matriz.

Lema 3.4.1 *Considere o sistema linear não-homogêneo $Tx = b$, em que $T : \mathbb{R}^n \to \mathbb{R}^m$ é representada pela matriz $A = (a_{ij})$. Então são equivalentes:*

(i) Existe solução x para $Tx = b$;

(ii) O vetor b é combinação linear das colunas de A.

Demonstração: Basta notar que o sistema $Tx = b$ é equivalente à equação

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

□

Em outras palavras, acabamos de mostrar que $< \mathcal{C} >$ é o subespaço $\mathcal{I}m\, T$, imagem da aplicação linear T.

Definição 3.4.2 *Se $A = (a_{ij})$ é uma matriz $m \times n$, definimos a matriz transposta de A como a matriz A^T de ordem $n \times m$ cujo elemento ij é a_{ji}.*

Em outras palavras, se A é a matriz dada por (3.7), então

$$A^T = \begin{pmatrix} a_{11} & \ldots & a_{m1} \\ \vdots & \ldots & \vdots \\ a_{1n} & \ldots & a_{mn} \end{pmatrix}.$$

Assim, as colunas da matriz A^T são justamente as linhas da matriz A. Como conseqüência imediata do lema 3.4.1 temos que

$$< \mathcal{L} >= \mathcal{I}m\ A^T. \tag{3.8}$$

Se S é a aplicação linear representada pela matriz A (com relação às bases canônicas do $\mathbb{R}^n$ e $\mathbb{R}^m$), então $< \mathcal{L} >$ é a imagem da aplicação linear S^T (que é chamada **transposta** da aplicação linear S e representada pela matriz A^T).

Vamos agora relacionar as dimensões dos espaços $< \mathcal{C} >$ e $< \mathcal{L} >$ de uma matriz A. Mostraremos que esses espaços têm a mesma dimensão; isso é um fato notável, pois eles são subespaços de espaços vetoriais diferentes!

Teorema 3.4.3 *Dada uma matriz $m \times n$, seu espaço-linha tem a mesma dimensão de seu espaço-coluna.*

Demonstração: Suponhamos que os vetores

$$b_1 = (b_{11}, b_{12}, \ldots, b_{1n}),\ b_2 = (b_{21}, b_{22}, \ldots, b_{2n}),\ \ldots,\ b_r = (b_{r1}, b_{r2}, \ldots, b_{rn})$$

formem uma base do espaço-linha da matriz A. Então cada linha ℓ_i de A é combinação linear desses elementos:

$$\begin{array}{rcl} \ell_1 & = & \lambda_{11}b_1 + \ldots + \lambda_{1r}b_r \\ \ell_2 & = & \lambda_{21}b_1 + \ldots + \lambda_{2r}b_r \\ \vdots & = & \vdots \\ \ell_m & = & \lambda_{m1}b_1 + \ldots + \lambda_{mr}b_r \end{array}$$

Igualando a componente i de cada uma das equações acima, obtemos

$$\begin{array}{rcl} a_{1i} & = & \lambda_{11}b_{1i} + \lambda_{12}b_{2i} + \ldots + \lambda_{1r}b_{ri} \\ a_{2i} & = & \lambda_{21}b_{1i} + \lambda_{22}b_{2i} + \ldots + \lambda_{2r}b_{ri} \\ \vdots & = & \vdots \\ a_{mi} & = & \lambda_{m1}b_{1i} + \lambda_{m2}b_{2i} + \ldots + \lambda_{mr}b_{ri}. \end{array}$$

Quer dizer,

$$\begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{pmatrix} = b_{1i} \begin{pmatrix} \lambda_{11} \\ \lambda_{21} \\ \vdots \\ \lambda_{m1} \end{pmatrix} + b_{2i} \begin{pmatrix} \lambda_{12} \\ \lambda_{22} \\ \vdots \\ \lambda_{m2} \end{pmatrix} + \ldots + b_{ri} \begin{pmatrix} \lambda_{1r} \\ \lambda_{2r} \\ \vdots \\ \lambda_{mr} \end{pmatrix},$$

mostrando que as colunas de A são combinações lineares dos r vetores

$$\begin{pmatrix} \lambda_{11} \\ \lambda_{21} \\ \vdots \\ \lambda_{m1} \end{pmatrix}, \ldots, \begin{pmatrix} \lambda_{1r} \\ \lambda_{2r} \\ \vdots \\ \lambda_{mr} \end{pmatrix}.$$

Isso quer dizer que o espaço-coluna tem dimensão, no máximo, igual a r, ou seja,

$$\dim < \mathcal{C} > \leq \dim < \mathcal{L} > .$$

Procedendo da mesma maneira com relação a uma base do espaço-coluna, mostramos que

$$\dim < \mathcal{L} > \leq \dim < \mathcal{C} > .$$

Assim, essas duas dimensões são iguais.

$\square$

Temos então a seguinte conseqüência imediata:

Corolário 3.4.4 *Seja A uma matriz $m \times n$. Então* $\dim \mathcal{I}m\, A = \dim \mathcal{I}m\, A^T$.

Se denotamos $r := \dim \mathcal{I}m\, A = \dim \mathcal{I}m\, A^T$, a aplicação do teorema do núcleo e da imagem garante:

$$\dim \ker A = n - r \quad \text{e} \quad \dim \ker A^T = m - r.$$

Assim,

Corolário 3.4.5 *Seja A uma matriz $n \times n$. Então*

$$\dim \ker A = \dim \ker A^T.$$

Esse resultado só vale para matrizes quadradas.

3.5 Aplicações lineares e matrizes II

Na primeira seção desse capítulo mostramos como associar a cada aplicação linear $T : \mathbb{R}^n \to \mathbb{R}^m$ uma matriz $A = (a_{ij})$ que representa T com relação às bases canônicas do $\mathbb{R}^n$ e $\mathbb{R}^m$. Note que o mesmo procedimento associa a cada aplicação linear $T : \mathbb{C}^n \to \mathbb{C}^m$ uma matriz $A = (a_{ij})$ que representa T com relação às bases canônicas do $\mathbb{C}^n$ e $\mathbb{C}^m$. Mostraremos agora que a mesma associação entre aplicações lineares e matrizes é válida para o caso de uma aplicação linear $T : X \to Y$ **entre espaços vetoriais de dimensão finita** X e Y.

A principal diferença, nesse caso, consiste em não termos uma escolha "natural" para bases nos espaços X e Y. Suponhamos que $\dim X = n$ e $\dim Y = m$. Escolhendo uma base arbitrária $\mathcal{B} = \{x_1, \ldots, x_n\}$ do espaço X e escrevendo $x = \lambda_1 x_1 + \ldots + \lambda_n x_n$, a aplicação $B : X \to \mathbb{K}^n$ definida por $Bx = (\lambda_1, \ldots, \lambda_n) = \lambda_1 e_1 + \ldots + \lambda_n e_n$ é um isomorfismo entre X e $\mathbb{K}^n$. Da mesma forma, ao se escolher uma base $\mathcal{C} = \{y_1, \ldots, y_m\}$ no espaço Y, se obtém um isomorfismo C entre Y e $\mathbb{K}^m$. Temos assim o seguinte diagrama:

$$\begin{array}{ccc} & T & \\ X & \longrightarrow & Y \\ B \downarrow & & \downarrow C \\ \mathbb{K}^n & \longrightarrow & \mathbb{K}^m \\ & T_{\mathbb{K}} & \end{array} . \tag{3.9}$$

A aplicação linear $T_{\mathbb{K}}$ é definida como composta de aplicações lineares (estamos usando a notação de composta para enfatizar)

$$T_{\mathbb{K}} = C \circ T \circ B^{-1}$$

e é representada por uma matriz A, de acordo como o que vimos na primeira seção desse capítulo. É usual chamar a matriz A de **representação da aplicação linear** T com respeito às bases $\mathcal{B}$ e $\mathcal{C}$ (dos espaços X e Y, respectivamente) e denotar $A = T_{\mathcal{B}}^{\mathcal{C}}$. Temos, assim, uma identificação entre a aplicação linear T e a matriz $A = T_{\mathcal{B}}^{\mathcal{C}}$. Com essa identificação, o diagrama (3.9) pode ser condensado:

$$X, \mathcal{B} \xrightarrow{T_{\mathcal{B}}^{\mathcal{C}}} Y, \mathcal{C} \tag{3.10}$$

(estamos enfatizando, na expressão dos espaços X e Y, as bases que produziram a matriz $T_{\mathcal{B}}^{\mathcal{C}}$). É fácil verificar (veja exercício 8) que a inversa T^{-1}, quando existe, é representada pela matriz $A^{-1} = [T_{\mathcal{B}}^{\mathcal{C}}]^{-1}$, chamada **matriz inversa** da matriz A.

Note que, em particular, esse mesmo raciocínio pode ser empregado no caso de uma aplicação linear $T : \mathbb{R}^n \to \mathbb{R}^m$, se escolhermos bases arbitrárias em $\mathbb{R}^n$ e $\mathbb{R}^m$.

Associamos assim a cada aplicações linear $T : X \to Y$ uma matriz, cuja expressão depende dos isomorfismos entre X e $\mathbb{K}^n$ e Y e $\mathbb{K}^m$. Esses, por sua vez, dependem das bases consideradas nos espaços X e Y. Uma vez que cada escolha de base em X produz um isomorfismo diferente entre X e $\mathbb{K}^n$ e o mesmo acontece com Y e $\mathbb{K}^m$, vemos que existem muitas maneiras distintas de representar uma transformação linear por meio de uma matriz. Como se relacionam essas diferentes matrizes que representam a aplicação linear T?

Seja, portanto, uma outra representação $T_{\bar{\mathcal{B}}}^{\bar{\mathcal{C}}}$, relativa às bases $\bar{\mathcal{B}}$ de X e $\bar{\mathcal{C}}$ de Y. Consideremos a aplicação linear que leva as coordenadas de x na base $\mathcal{B}$ nas suas coordenadas na base $\bar{\mathcal{B}}$. Essa aplicação é um isomorfismo e é representada por uma matriz, como acabamos de sintetizar no diagrama 3.10. Essa matriz é denotada $P_{\mathcal{B}}^{\bar{\mathcal{B}}} : X \to X$ e chamada **matriz mudança** da base $\mathcal{B}$ para a base $\bar{\mathcal{B}}$ (no exercício 7 se pede para mostrar que essa matriz corresponde à aplicação "identidade" entre X com a base $\mathcal{B}$ e X com a base $\bar{\mathcal{B}}$). Da mesma forma, temos o isomorfismo $Q_{\mathcal{C}}^{\bar{\mathcal{C}}}$, mudança da base $\mathcal{C}$ para a base $\bar{\mathcal{C}}$. Temos, assim, o diagrama

$$\begin{array}{ccc} X, \mathcal{B} & \xrightarrow{T_{\mathcal{B}}^{\mathcal{C}}} & Y, \mathcal{C} \\ P_{\mathcal{B}}^{\bar{\mathcal{B}}} \downarrow & & \downarrow Q_{\mathcal{C}}^{\bar{\mathcal{C}}} \\ X, \bar{\mathcal{B}} & \xrightarrow[T_{\bar{\mathcal{B}}}^{\bar{\mathcal{C}}}]{} & Y, \bar{\mathcal{C}} \end{array} .$$

Esse diagrama, cujos componentes são matrizes, nos mostra que

$$T_{\mathcal{B}}^{\mathcal{C}} = [Q_{\mathcal{C}}^{\bar{\mathcal{C}}}]^{-1} T_{\bar{\mathcal{B}}}^{\bar{\mathcal{C}}} P_{\mathcal{B}}^{\bar{\mathcal{B}}}.$$

Note que $[Q_{\mathcal{C}}^{\bar{\mathcal{C}}}]^{-1} = Q_{\bar{\mathcal{C}}}^{\mathcal{C}}$ (veja exercício 9), de modo que

$$T_{\mathcal{B}}^{\mathcal{C}} = Q_{\bar{\mathcal{C}}}^{\mathcal{C}} T_{\bar{\mathcal{B}}}^{\bar{\mathcal{C}}} P_{\mathcal{B}}^{\bar{\mathcal{B}}}.$$

O caso em que os espaços X e Y são iguais permite que se tome a mesma base nos dois espaços. Nesse caso, denotamos $T_{\mathcal{B}}^{\mathcal{B}}$ por $T_{\mathcal{B}}$, que é chamada representação de T na base $\mathcal{B}$. A relação entre $T_{\mathcal{B}}$ e $T_{\bar{\mathcal{B}}}$ é dada por

$$T_{\bar{\mathcal{B}}} = [P_{\bar{\mathcal{B}}}^{\mathcal{B}}]^{-1} T_{\mathcal{B}} P_{\bar{\mathcal{B}}}^{\mathcal{B}} = P_{\mathcal{B}}^{\bar{\mathcal{B}}} T_{\mathcal{B}} P_{\bar{\mathcal{B}}}^{\mathcal{B}},$$

para qualquer outra base $\bar{\mathcal{B}}$ de X.

Exemplo 3.5.1 Considere a aplicação linear $T : \mathbb{R}^2 \to \mathbb{R}^2$ definida por

$$T(x, y) = (4x - 2y, 2x + y).$$

Para simplificarmos a notação nesse exemplo, escreveremos os nossos vetores indiferentemente como linhas ou colunas.

Sejam $\mathcal{B}$ a base do $\mathbb{R}^2$ formada pelos vetores $v_1 = (1, 1)$ e $v_2 = (-1, 0)$. Vamos achar a matriz que representa T com relação à base $\mathcal{B}$. Quer dizer, estamos utilizando a mesma base no domínio e na imagem e procuramos a matriz $T_{\mathcal{B}}$. Para isso, calculamos

$$T(v_1) = (2, 3) = 3(1, 1) + (-1, 0) = 3v_1 + v_2.$$

Note que escrevemos a imagem de $T(v_1)$ na base $\mathcal{B}$, utilizada também no contradomínio. De acordo com a notação introduzida na definição 1.3.8, temos

$$[T(v_1)]_{\mathcal{B}} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

Da mesma forma, $T(v_2) = (-4, -2) = -2(1, 1) + 2(-1, 0) = -2v_1 + 2v_2$ e, portanto,

$$[T(v_2)]_{\mathcal{B}} = \begin{pmatrix} -2 \\ 2 \end{pmatrix}.$$

Assim,

$$T_{\mathcal{B}} = \begin{pmatrix} 3 & -2 \\ 1 & 2 \end{pmatrix}.$$

As colunas de $T_{\mathcal{B}}$ são as imagens dos vetores da base $\mathcal{B}$, escritas na própria base $\mathcal{B}$ utilizada, nesse caso, também no contradomínio.

Se quisermos calcular a imagem do vetor $(1, 2) = 1e_1 + 2e_2$ utilizando a matriz $T_{\mathcal{B}}$, primeiro expressamos esse vetor na base $\mathcal{B}$:

$$(1, 2) = 2(1, 1) + 1(-1, 0) = 2v_1 + v_2.$$

Calculando

$$T_{\mathcal{B}} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 & -2 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 4 \end{pmatrix},$$

obtemos a "resposta" na base $\mathcal{B}$. Se quisermos a resposta na base canônica, precisamos escrever o resultado obtido nessa base:

$$4(1, 1) + 4(-1, 0) = (0, 4) = 0e_1 + 4e_2,$$

que é o mesmo que calcular diretamente $T(1, 2)$ utilizando a expressão $T(x, y) = (4x-2y, 2x+y)$.

Para entendermos melhor a estrutura desse exemplo, temos o seguinte diagrama

$$\begin{array}{ccc} & T_{\mathcal{E}} & \\ \mathbb{R}^2, \mathcal{E} & \longrightarrow & \mathbb{R}^2, \mathcal{E} \\ P_{\mathcal{E}}^{\mathcal{B}} \downarrow & & \downarrow P_{\mathcal{E}}^{\mathcal{B}} \\ \mathbb{R}^2, \mathcal{B} & \longrightarrow & \mathbb{R}^2, \mathcal{B} \\ & T_{\mathcal{B}} & \end{array} .$$

Aqui, $T_{\mathcal{E}}$ é a representação "natural" da transformação $T(x,y) = (4x - 2y, 2x + y)$. Isso é, a matriz cujas **colunas** são, respectivamente, $T(1,0) = (4\ 2)$ e $T(0,1) = (-2\ 1)$.)

A matriz $T_{\mathcal{B}}$ é a matriz obtida no exemplo. A matriz $P_{\mathcal{E}}^{\mathcal{B}}$ é a matriz mudança da base $\mathcal{E}$ para a base $\mathcal{B}$. Ela é obtida pelo mesmo método (veja exercício 7): escrevemos a imagem dos vetores e_1, e_2 pela aplicação identidade na base $\mathcal{B}$. Temos

$$(1,0) = 0(1,1) - 1(-1,0) = 0v_1 - v_2 \quad \text{e} \quad (0,1) = 1(1,1) + 1(-1,0) = 1v_1 + 1v_2.$$

A matriz $P_{\mathcal{E}}^{\mathcal{B}}$ é, então,

$$P_{\mathcal{E}}^{\mathcal{B}} = \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}.$$

O diagrama anterior garante que

$$T_{\mathcal{E}} = [P_{\mathcal{E}}^{\mathcal{B}}]^{-1} T_{\mathcal{B}} P_{\mathcal{E}}^{\mathcal{B}},$$

ou seja,

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

Se calcularmos a inversa da matriz $P_{\mathcal{E}}^{\mathcal{B}}$, verificaremos esse fato. Entretanto, é fácil obter $P_{\mathcal{B}}^{\mathcal{E}}$. Essa matriz tem como colunas a expressão dos vetores v_1 e v_2 na base canônica. Assim, é claro que

$$P_{\mathcal{B}}^{\mathcal{E}} = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}.$$

Verifique que $P_{\mathcal{B}}^{\mathcal{E}} = [P_{\mathcal{E}}^{\mathcal{B}}]^{-1}$. ◀

3.6 A transposta de uma aplicação linear

Existe uma maneira intrínseca de se definir a aplicação transposta T^T de um operador linear T. (No caso de aplicações lineares se denota a transposta T^T também por T', o que faremos a seguir).

Para isso, sejam $T : X \to Y$ uma aplicação linear entre os espaços X e Y e $\ell \in Y'$, isto é, $\ell : Y \to \mathbb{K}$ é linear. Então o produto dessas aplicações (isto é, a composta) $\ell T : X \to \mathbb{K}$ é um elemento do dual X'.

$$\begin{array}{ccc} & T & \\ X & \longrightarrow & Y \\ & \searrow & \downarrow \ell \\ & m_\ell & \mathbb{K} \end{array}$$

Estamos denotando (provisoriamente) $m_\ell(x) = \ell(Tx)$. Note que, variando $\ell \in Y'$, obtemos diferentes aplicações $m \in X'$. Consideremos então $T' : Y' \to X'$ definida por

$$T'(\ell) = \ell(Tx) = m_\ell(x).$$

Afirmamos que T' é linear. De fato,

$$T'(\ell_1 + \lambda\ell_2) = (\ell_1 + \lambda\ell_2)(Tx) = \ell_1(Tx) + \lambda\ell_2(Tx) = T'(\ell_1) + \lambda T'(\ell_2),$$

para quaisquer $\ell_1, \ell_2 \in Y'$ e $\lambda \in \mathbb{K}$. Desse modo, a aplicação T' é definida como uma aplicação linear definida no espaço dual Y' e tomando valores no espaço dual X'.

Vamos agora introduzir uma nova notação para a avaliação de um elemento do dual em um ponto do espaço: até agora estamos denotando, se $\ell : Z' \to \mathbb{K}$ e $z \in Z$, $\ell(z)$. Também denotaremos $\ell(z)$ por

$$\langle \ell, z \rangle.$$

Abandonaremos a notação provisória m_ℓ e usaremos a notação $T'\ell$. Assim, por definição,

$$\langle T'\ell, x \rangle = \langle \ell, Tx \rangle$$

ou, o que é o mesmo

$$T'\ell = \ell T. \tag{3.11}$$

Nosso próximo objetivo é caracterizar a aplicação T' para o caso de $T : \mathbb{R}^n \to \mathbb{R}^m$. Veremos que podemos representar T' (a aplicação transposta) por uma matriz, que é justamente a transposta da matriz que representa T com relação às bases canônicas do $\mathbb{R}^n$ e $\mathbb{R}^m$.

O lado direito de (3.11) tem interpretação imediata: como $\ell \in (\mathbb{R}^m)'$, ℓ é dada por uma matriz linha, de modo que

$$\ell T = (c_1 \ \ldots \ c_m) \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ldots & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix}.$$

Se quisermos interpretar T' como uma matriz, então devemos identificar $(\mathbb{R}^m)'$ com $\mathbb{R}^m$ e $(\mathbb{R}^n)'$ com $\mathbb{R}^n$. Assim $T' : (\mathbb{R}^m)' \to (\mathbb{R}^n)'$ passa a ser vista como uma aplicação $T : \mathbb{R}^m \to \mathbb{R}^n$. O vetor coluna $\ell \in \mathbb{R}^m$, quando aplicado a T', satisfaz a igualdade $T'\ell = \ell T$, ou seja, se $B = (b_{ij})$ é a representação matricial de T' (com relação às bases canônicas do $\mathbb{R}^m$ e $\mathbb{R}^n$), então

$$T' \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix} = \begin{pmatrix} b_{11} & \ldots & b_{1m} \\ \vdots & \ldots & \vdots \\ b_{n1} & \ldots & a_{nm} \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix} = (c_1 \ \ldots \ c_m) \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ldots & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix}.$$

A segunda igualdade acima mostra que $B = (b_{ij})$ deve satisfazer $b_{ij} = a_{ji}$, como se verifica mediante escolha adequada de $c_1, \ldots, c_m$. Mas então $B = A^T$, como antes definido.

3.7 Exercícios

1. Represente matricialmente a base dual da base $\{e_1, \ldots, e_n\}$ do $\mathbb{R}^n$.

2. Mostre a proposição 3.3.8 utilizando o isomorfismo T_q definido na primeira demonstração do teorema do núcleo e da imagem.

3. Seja $X = W_1 \oplus W_2$ e $x = w_1 + w_2$, com $w_i \in W_i$. Mostre que $\Pi : X \to W_1$, definida por $\Pi x = w_1$, é uma aplicação linear. Seja $\pi : X \to X$ uma aplicação linear tal que $\pi^2 = \pi$ (uma tal aplicação linear é chamada **projeção**). Mostre que $X = \ker \pi \oplus \mathcal{I}m\ \pi$. Mostre também que Π (definida acima) é uma projeção.

4. Sejam X um espaço vetorial e $Y \subset X$ um subespaço. Mostre que $\pi : X \to X/Y$ definida por $\pi(x) = x + Y = [x]$ é uma aplicação linear.

5. Sejam X e Y espaços vetoriais e $\mathcal{B}$ uma base de X (mesmo que X tenha dimensão infinita). Façamos corresponder, de maneira arbitrária, um vetor $y_x \in Y$ a cada elemento $x \in \mathcal{B}$. Mostre que existe uma única transformação linear $T : X \to Y$ tal que $Tx = y_x$ para todo $x \in \mathcal{B}$. (Note que, em particular, isso implica que uma transformação linear $T : \mathbb{K}^n \to \mathbb{K}^m$ fica completamente determinada pela imagem que ela assume em qualquer base do $\mathbb{K}^n$). Mostre então que uma transformação linear $T : X \to Y$ é injetiva se, e somente se, leva vetores linearmente independentes em vetores linearmente independentes.

6. Sejam X e Y espaços vetoriais com a mesma dimensão. Suponhamos que, para as aplicações linear $T : X \to Y$ e $S : Y \to X$, seja verdadeiro $ST = I$, a identidade em X. Mostre que $S = T^{-1}$.

7. Verifique que a matriz $P_{\mathcal{B}}^{\bar{\mathcal{B}}} : X \to X$ corresponde à representação matricial da aplicação identidade $I : X \to X$, com relação às bases $\mathcal{B}$ e $\bar{\mathcal{B}}$.

8. Seja $T : X \to Y$ uma aplicação linear invertível representada, com relação às bases $\mathcal{B}$ e $\mathcal{C}$ dos espaços X e Y, respectivamente, pela matriz $T_{\mathcal{B}}^{\mathcal{C}}$. Mostre que a aplicação inversa T^{-1} é representada pela matriz $[T_{\mathcal{B}}^{\mathcal{C}}]^{-1}$.

9. Seja $P_{\mathcal{B}}^{\bar{\mathcal{B}}}$ a matriz mudança da base $\mathcal{B}$ para a base $\bar{\mathcal{B}}$. Mostre que $(P_{\mathcal{B}}^{\bar{\mathcal{B}}})^{-1} = P_{\bar{\mathcal{B}}}^{\mathcal{B}}$.

10. Seja V um espaço vetorial de dimensão finita sobre $\mathbb{K}$. Definimos, para $v, w \in V$, $v \equiv w$ se existe uma transformação linear invertível $T : V \to V$ tal que $Tv = w$. Mostre que assim está definida uma relação de equivalência. Mostre também que essa relação de equivalência possui apenas duas classes: uma formada apenas pelo elemento $0 \in V$ e a outra formada por todos os outros vetores de V.

11. Considere os polinômios $p_1(t) = 7t^5 + 6t^2$, $p_2(t) = 1 + t$ no espaço $\mathcal{P}_6$ de todos os polinômios de grau menor que 6.

    (a) Se $S = \{p_1, p_2\}$, descreva $< S >$;

    (b) ache uma base $\mathcal{B}$ de $\mathcal{P}_6$ que completa o conjunto linearmente independente S;

    (c) determine a representação de cada um dos vetores de $\mathcal{B}$ nessa base;

    (d) determine a representação de $q \in \mathcal{P}_6$ em termos da base $\mathcal{B}$.

12. Seja $\mathcal{P}$ o espaço de todos os polinômios na variável t. Considere $T : \mathcal{P} \to \mathcal{P}_6$ definida da seguinte maneira: se $p \in \mathcal{P}$ então Tp é o polinômio em $\mathcal{P}_6$ cujos coeficientes de grau menor que 6 são iguais aos coeficientes de p. Mostre que T é linear. Ache uma base para $\mathcal{I}m\,T$ e $\ker T$. O teorema do núcleo e da imagem se aplica? Justifique.

13. Se

$$M = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

defina $T : \mathbb{M}_{2\times 2} \to \mathbb{M}_{2\times 3}$ por

$$T(M) = \begin{pmatrix} a_{12} & a_{11} - a_{12} & a_{21} - a_{12} \\ a_{22} & a_{21} - a_{11} & a_{22} + a_{21} \end{pmatrix}.$$

Sejam

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right\},$$
$$\mathcal{B}' = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\},$$

$$\begin{aligned} \mathcal{C} &= \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \right. \\ &\qquad \left. \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \right\} \end{aligned}$$

$$\begin{aligned} \mathcal{C}' &= \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \right. \\ &\qquad \left. \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right\}. \end{aligned}$$

(a) Mostre que $T : \mathbb{M}_{2\times 2} \to \mathbb{M}_{2\times 3}$ é linear;

(b) mostre que $\mathcal{B}$ e $\mathcal{B}'$ são bases de $\mathbb{M}_{2\times 2}$, enquanto $\mathcal{C}$ e $\mathcal{C}'$ são bases de $\mathbb{M}_{2\times 3}$;

(c) ache a representação matricial de T relativa às bases $\mathcal{B}$ e $\mathcal{C}$, bem como a relativa às bases $\mathcal{B}'$ e $\mathcal{C}'$;

(d) ache a relação entre essas matrizes;

(e) obtenha bases para $\ker T$ e $\mathcal{I}m\, T$.

14. Sejam $T(x, y, x) = (x + y + z, y + z, x)$ e $\mathcal{B} = \{(1, 0, 1), (0, 1, 0), (-1, 0, 1)\}$. Então:

    (a) ache a matriz $T_{\mathcal{B}}$;

    (b) usando a matriz acima, especifique uma base para $\ker T$ e $\mathcal{I}m\, T$;

    (c) calcule $T(1, 1, 1)$ utilizando a representação matricial calculada em (a).

15. A definição dos espaços $\ker T$ e $\mathcal{I}m\, T$ de uma aplicação linear $T : X \to Y$ independe (da existência) de bases nesses espaços. Contudo, se A é uma matriz que representa uma transformação linear, tanto $\ker A$ como $\mathcal{I}m\, A$ dependem das bases consideradas no domínio e no contradomínio. Explique.

16. Sejam X um espaço vetorial de dimensão finita e $T : X \to X$ uma aplicação linear. Mostre que
$$X = \ker T \oplus \mathcal{I}m\, T$$
se, e somente se, $\ker T = \ker T^2$.

17. Justifique[2] o algoritmo utilizado para se obter a inversa de uma matriz quadrada A.

18. Sejam A, B matrizes quadradas invertíveis. Mostre que $(AB)^{-1} = B^{-1}A^{-1}$.

19. Seja $A = (a_1\ a_2\ \ldots\ a_n)$ e B uma matriz cuja j-ésima coluna é $b_j = (b_{1j}\ b_{2j}\ \cdots b_{nj})^T$. Se está definido o produto AB, mostre que a j-ésima coluna de AB é dada por

$$Ab_j = b_{1j}a_1 + \ldots + b_{nj}a_n.$$

20. Se V é um espaço vetorial de dimensão finita n e $W_1, \ldots, W_n$ são subespaços de V tais que $V = W_1 \oplus \cdots \oplus W_n$, mostre que $\dim W_i = 1$.

    Seja agora $T : V \to V$ uma transformação linear e $V = W_1 \oplus \cdots \oplus W_k$ (os subespaços W_i não precisam ter dimensão igual a 1). Suponhamos que $T(W_i) \subset W_i$ para $i \in \{1, \ldots, k\}$ (dizemos que os subespaços W_i são **invariantes** por T. Se $\mathcal{B}_i$ for uma base de W_i, mostre que $\mathcal{B} = \bigcup_{i=1}^{k} \mathcal{B}_i$ é uma base de V. Ache $T_{\mathcal{B}}$, a representação de T na base $\mathcal{B}$ em termos de $T_{\mathcal{B}_i}$, a representação de $T : W_i \to W_i$ na base $\mathcal{B}_i$.

21. Sejam $A, B \in \mathbb{M}_{n\times n}$, o espaço das matrizes $n \times n$ com coeficientes em $\mathbb{K}$. Definimos $A \sim B$ se existe uma matriz invertível $P \in \mathbb{M}_{n\times n}$ tal que $B = P^{-1}AP$. Mostre que $A \sim B$ é uma relação de equivalência. Esboce um diagrama que representa essa relação de equivalência. É usual dizer então que A e B são iguais, a menos da uma mudança de base. Você consegue dar um sentido para essa frase?

[2] Para esse exercício é necessário o conhecimento do conceito de matrizes elementares. Veja a seção 8.1.

Capítulo 4

Determinantes

4.1 Permutações

Definição 4.1.1 *Seja $S = \{1, 2, \ldots, n\}$ ou, mais geralmente, um conjunto $\{x_1, \ldots, x_n\}$ com n elementos distintos. Uma* **permutação** *é uma aplicação sobrejetiva $p : S \to S$.*

É claro que p é, necessariamente, injetiva. Assim, permutações podem ser compostas e têm inversa. Denotaremos p^0 a permutação identidade, $q \circ p = qp$ a composta de duas permutações e definimos, para $k \in \mathbb{N}^*$, $p^k = pp^{k-1}$. Definimos, para $k \in \mathbb{Z}$, $k < 0$, $p^k = (p^{-1})^{-k}$.

Existem várias notações para uma permutação $p : S \to S$. Em geral escrevemos $p(i) = p_i$ (ou $p(x_i) = p_i$) e denotamos

$$p(1, \ldots, n) = \begin{pmatrix} 1 & 2 & \ldots & n \\ p_1 & p_2 & \ldots & p_n \end{pmatrix} \quad \text{ou} \quad p = \frac{1\ 2\ \ldots\ n}{p_1 p_2 \ldots p_n}.$$

Exemplo 4.1.2 Considere a permutação

$$\frac{1\ 2\ 3\ 4}{2\ 4\ 1\ 3}.$$

Então

$$p^2 = \frac{1\ 2\ 3\ 4}{4\ 3\ 2\ 1}, \quad p^{-1} = \frac{1\ 2\ 3\ 4}{3\ 1\ 4\ 2}, \quad p^3 = \frac{1\ 2\ 3\ 4}{3\ 1\ 4\ 2}, \quad p^4 = \frac{1\ 2\ 3\ 4}{1\ 2\ 3\ 4} = p^0.$$

◀

Definição 4.1.3 *Seja $p : S \to S$ uma permutação. Dados $a, b \in S$, definimos $a \sim b \mod p$ se existe $i \in \mathbb{Z}$ tal que $b = p^i(a)$.*

Isso estabelece uma relação de equivalência[1] em S. De fato:

(*i*) $a \sim a \mod p$, pois $a = p^0(a)$;

(*ii*) $a \sim b \mod p$ implica $b \sim a \mod p$, pois $b = p^i(a)$ implica $p^{-i}(b) = p^{-i}(p^i(a)) = a$;

[1]Veja exercício 23 do capítulo 1.

(*iii*) $a \sim b \mod p$ e $b \sim c \mod p$ implica $a \sim c \mod p$, pois $b = p^i(a)$ e $c = p^j(b)$ e, portanto, $c = p^j(p^i(a)) = p^{j+i}(a)$.

Definição 4.1.4 *Chamamos de* **órbita** *de a a classe de equivalência a que pertence o elemento a. Assim, a órbita de a consiste de todos os elementos $p^i(a)$, $i \in \mathbb{Z}$.*

Entretanto, $p^i(a) \in S$ para todo i. Assim, existe um menor inteiro positivo k (que depende do elemento a) tal que $p^k(a) = a$. O número $k = k_a$ é chamado **ordem** do elemento a.

Definição 4.1.5 *O* **ciclo** *de a é o conjunto ordenado $\{a, p(a), p^2(a), \ldots, p^{k_a-1}(a)\}$.*

Veremos como identificar o ciclo de a com uma permutação $\sigma : S \to S$. Primeiramente mostraremos

Lema 4.1.6 *Todos os elementos da órbita de a estão presentes no ciclo de a. Os elementos do ciclo de a são distintos.*

Demonstração: Consideremos $s \in \mathbb{Z}$ e $p^s(a)$. Seja k a ordem do elemento a. Então existem inteiros m e r tais que $s = mk + r$, com $0 \leq r < k$ (divisão euclidiana). Mas então

$$p^s(a) = p^{mk+r}(a) = p^{mk}(p^r(a)) = p^r(a),$$

mostrando a primeira afirmação (veja exercício 1).

Se fosse $p^i = p^j$ para $0 \leq i < j \leq (k-1)$, então $a = p^0 = p^{j-i}$, com $j - i < k$, contradizendo a definição de k.

□

Se conhecemos todos os ciclos de uma permutação p, conhecemos a imagem de todos os elementos de p. A cada ciclo corresponde uma permutação de S. De fato, para cada $a \in S$ e $0 \leq i \leq k_a - 1$, considere a permutação que envia $p^i(a)$ em $p^{i+1}(a)$, os elementos que não pertencem ao ciclo de a permanecendo inalterados. É usual chamar de **ciclo** à permutação assim obtida (e não mais ao conjunto obtido através das órbitas). Daremos um exemplo ilustrando essas idéias:

Exemplo 4.1.7 Seja $S = \{1, 2, \ldots, 6\}$. Consideremos $p : S \to S$ definida por $(2\ 1\ 3\ 5\ 6\ 4)$. Então $1 = p^0(1)$, $2 = p^1(1)$, $p^2(1) = p(p(1)) = p(2) = 1$. Assim, o ciclo de 1 é $\{1, 2\}$. O ciclo de 3 consiste apenas do 3; o ciclo de 4 consiste dos elementos 4, $5 = p(4)$, $6 = p^2(4) = p(5)$, pois $p^3(4) = p(p^2(4)) = p(6) = 4$. Note que $p(1) = 2$, $p(2) = 1$, $p(3) = 3$, $p(4) = 5$, $p(5) = 6$, $p(6) = 4$. Assim, conhecemos a imagem por p de todos os elementos de S.

Ao conjunto $\{1, 2\}$ corresponde a permutação $p_1 : S \to S$ tal que $p_1(1) = 2$, $p_1(1) = 1$, os outros elementos permanecendo fixos. Vamos denotar tal permutação por $(1\ 2)$. (Note que, para um conjunto S fixo, essa notação não ocasiona ambigüidade). Da mesma forma, seja p_2 a permutação definida por $p_2(4) = 5$, $p_2(5) = 6$, $p_2(6) = 4$, os outros elementos permanecendo fixos; vamos denotar p_2 por $(4\ 5\ 6)$. Verifique que $p = p_1p_2 = p_2p_1$ (ao conjunto $\{3\}$ corresponde a permutação identidade). As permutações p_1, p_2 e p_3 são os ciclos de p. É usual desprezar o ciclo identidade p_3 e dizer que os ciclos de p são p_1 e p_2. Note que os ciclos de p são disjuntos, pois foram gerados por classes de uma relação de equivalência. ◀

O que aconteceu no exemplo acima é um fato geral:

Lema 4.1.8 *Toda permutação é o produto (quer dizer, a composição) de seus ciclos.*

Demonstração: Seja $p : S \to S$ e $s \in S$. Sejam $\rho_1, \ldots, \rho_j$ os ciclos de p. Como esses ciclos são gerados por classes de equivalência, existe $i \in \{1, \ldots, j\}$ tal que $s \in p_i$. De acordo com o exercício 1, temos que

$$\rho_i = (s\ p(s)\ \ldots\ p^{k-1}(s))$$

em que k é a ordem de s. Mas o ciclo ρ_k afeta apenas os elementos que ele contém, os outros permanecendo fixos. Assim, $(\rho_1 \ldots \rho_i \ldots \rho_j)(s) = \rho_i(s) = p(s)$. Isso mostra o afirmado. Note que é irrelevante a ordem em que o produto de ciclos é tomado.

□

Definição 4.1.9 *Uma* **transposição** *é uma permutação $p : S \to S$ tal que existem dois elementos $i, j \in S$ (ou $a_i, a_j \in S$) com*

$$p(k) = k,\ \forall\ k \in S,\ k \neq i,\ k \neq j;\ p(i) = j \text{ e } p(j) = i.$$

Consideremos um ciclo $(a_1\ a_2\ \ldots a_m) = (a_1\ p(a_1)\ \ldots p^{m-1}(a_1))$. É fácil verificar que esse ciclo pode ser escrito como produto de transposições:

$$(a_1\ a_2\ \ldots a_m) = (a_1\ a_m) \ldots (a_1\ a_3)(a_1\ a_2).$$

Essa decomposição, entretanto, não é única. Por exemplo, podemos escrever

$$(1\ 2\ 3) = (1\ 3)(1\ 2) = (3\ 2)(3\ 1).$$

Vamos mostrar, entretanto, que o número de fatores numa decomposição de um ciclo como produto de transposições é sempre par ou sempre ímpar. Para simplificar a notação, denotaremos $S = \{x_1, \ldots, x_n\}$ ao invés de $\{1, \ldots, n\}$.

Definição 4.1.10 *Seja $S = \{x_1, \ldots, x_n\}$. Definimos o* **discriminante** *$P(x_1, \ldots, x_n)$ por*

$$P(x_1, \ldots, x_n) = \prod_{i<j} (x_i - x_j).$$

Exemplo 4.1.11 Consideremos $S = \{x_1, x_2, x_3, x_4\}$. Então

$$P(x_1, x_2, x_3, x_4) = (x_1 - x_2)(x_1 - x_3)(x_1 - x_4)(x_2 - x_3)(x_2 - x_4)(x_3 - x_4).$$

Note que $\prod_{i<j}(x_i - x_j)$ contém todos as combinações de termos $(x_i - x_j)$, com $i \neq j$. O discriminante apenas prescreve uma ordem para essas combinações: aquelas com $i < j$. ◀

Se $p : S \to S$ é uma permutação no conjunto $S = \{x_1, x_2, \ldots, x_n\}$, definimos

$$P_p(x_1, x_2, \ldots, x_n) = \prod_{i<j} (x_{p_i} - x_{p_j}).$$

Afirmamos que

$$P_p(x_1, x_2, \ldots, x_n) = \pm P(x_1, x_2, \ldots, x_n).$$

De fato, como p é uma permutação, todas as combinações $x_i - x_j$ constam de $P_p(x_1, \ldots, x_n)$. Entretanto, ou elas aparecem com $i < j$ ou com $j > i$. Isso mostra que elas podem diferir apenas em módulo.

Exemplo 4.1.12 Considere a permutação

$$p(x_1, x_2, x_3, x_4) = \left(\begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \\ x_4 & x_2 & x_3 & x_1 \end{array} \right).$$

Então

$$\begin{aligned} P_p(x_1, x_2, x_3, x_4) &= \prod_{i<j} (x_{p_i} - x_{p_j}) \\ &= (x_{p_1} - x_{p_2})(x_{p_1} - x_{p_3})(x_{p_1} - x_{p_4})(x_{p_2} - x_{p_3})(x_{p_2} - x_{p_4})(x_{p_3} - x_{p_4}) \\ &= (x_4 - x_2)(x_4 - x_3)(x_4 - x_1)(x_2 - x_3)(x_2 - x_1)(x_3 - x_1) \\ &= [-(x_2 - x_4)][-(x_3 - x_4)][-(x_1 - x_4)](x_2 - x_3)[-(x_1 - x_2)][-(x_1 - x_3)] \\ &= -(x_1 - x_2)(x_1 - x_3)(x_1 - x_4)(x_2 - x_3)(x_2 - x_4)(x_3 - x_4) \\ &= -P(x_1, \ldots, x_4). \end{aligned}$$

◀

Definição 4.1.13 *O* **sinal** $\sigma(p)$ (*ou* **paridade**) *de uma permutação p é definido por*

$$P_p(x_1, \ldots, x_n) = \sigma(p) P(x_1, \ldots, x_n).$$

Claramente temos $\sigma(p) = \pm 1$.

Note que se $\sigma(p) = 1$, então p altera um número par de termos $x_i - x_j$ na definição do discriminante e, se $\sigma(p) = -1$, p altera um número ímpar de termos. Assim, se τ é uma transposição, então $\sigma(\tau) = -1$ (justifique esse argumento no exercício 2). Além disso, temos

Lema 4.1.14 *Se p_1, p_2 são permutações, então*

$$\sigma(p_2 p_1) = \sigma(p_2)\sigma(p_1).$$

Demonstração: De fato, suponhamos que $\sigma(p_1) = -1 = \sigma(p_2)$. Então p_1 altera um número ímpar k_1 de termos de $\{x_1, \ldots, x_n\}$ e p_2 altera um número ímpar k_2 de termos. Escrevemos $k_2 = n_1 + n_2$, em que $n_1 \geq 0$ é a quantidade de termos alterados por p_2 dentre aqueles que já havia sido alterado por p_1 (esses ficam, portanto, inalterados). Assim, o total de termos alterados por $p_2 p_1 = p_2 \circ p_1$ é

$$k_1 - n_1 + n_2 = k_1 - n_1 + k_2 - n_1 = k_1 + k_2 - 2n_1.$$

Como $k_1 + k_2$ é par, temos que o total de termos alterados é par. O mesmo argumento se aplica a todos os outros casos.

□

É fácil agora notar que o número de transposições na decomposição de uma permutação é sempre par ou sempre ímpar. Suponhamos, por exemplo, que numa decomposição de p tenhamos encontrado um número par de transposições. Como $\sigma(\tau) = -1$ para toda transposição τ, concluímos que $\sigma(p) = 1$. Se fosse possível escrever p como um produto de um número ímpar de transposições, teríamos $\sigma(p) = -1$, absurdo. Em outras palavras, mostramos que se $p = \tau_k \circ \cdots \circ \tau_1$, então

$$\sigma(p) = (-1)^k. \tag{4.1}$$

4.2 Determinantes

Definição 4.2.1 *Sejam $a_1, a_2, \ldots, a_n$ pontos do $\mathbb{R}^n$. Definimos o* **determinante** $D(a_1, \ldots, a_n)$ *desses pontos como uma função*

$$\begin{array}{rccccccc} D : \mathbb{R}^n & \times & \cdots & \times & \mathbb{R}^n & \to & \mathbb{R} \\ (a_1 & , & \ldots & , & a_n) & \mapsto & D(a_1, \ldots, a_n) \end{array}$$

satisfazendo às seguintes propriedades:

(*i*) $D(a_1, \ldots, a_n) = 0$ *se* $a_i = a_j$ *para* $i \neq j$, $i, j \in \{1, \ldots, n\}$

(*ii*) $D(a_1, \ldots, a_n)$ *é uma aplicação n-linear, isto é, é uma aplicação linear de cada coordenada, as outras sendo mantidas fixas; em outras palavras, se todos os* a_i *com* $i \neq j$ *estão fixos,*

$$D(a_1, \ldots, \lambda x + y, \ldots, a_n) = \lambda D(a_1, \ldots, x, \ldots, a_n) + D(a_1, \ldots, y, \ldots, a_n).$$

(*iii*) $D(e_1, \ldots, e_n) = 1$, *em que* $\{e_1, \ldots, e_n\}$ *é a base canônica do* $\mathbb{R}^n$.

Se A é uma matriz $n \times n$ com vetores coluna $a_1, \ldots, a_n$, (quer dizer, $A = (a_1\ a_2\ \cdots\ a_n)$), definimos $\det A = D(a_1, a_2, \ldots, a_n)$.

Temos, como conseqüência imediata da definição do determinante,

Lema 4.2.2 *O determinante satisfaz às propriedades*

(*iv*) *D é uma aplicação linear* **alternada**, *isto é, se trocarmos a_i por a_j, então o valor do determinante é multiplicado por -1. Sendo mais preciso,*

$$D(a_1, \ldots, a_i, \ldots, a_j, \ldots, a_n) = -D(a_1, \ldots, a_j, \ldots, a_i, \ldots, a_n).$$

(*v*) *Se $a_1, \ldots, a_n$ são linearmente dependentes, então $D(a_1, \ldots, a_n) = 0$.*

Demonstração: Para mostrar (*iv*), uma vez que apenas os elementos a_i e a_j estão sendo trocados, indicaremos apenas essas coordenadas no determinante e denotaremos $a := a_i$ e $b := a_j$. Temos:

$$\begin{aligned} D(a, b) &= D(a, b) + D(a, a) = D(a, a + b) \\ &= D(a, a + b) - D(a + b, a + b) = -D(b, a + b) = -D(b, a) - D(b, b) \\ &= -D(b, a). \end{aligned}$$

Se $a_1, \ldots, a_n$ são linearmente dependentes, então um desses elementos pode ser escrito como combinação linear dos restantes. Vamos supor que $a_1 = \lambda_2 a_1 + \ldots + \lambda_n a_n$. Decorre então de (*iv*) que

$$\begin{aligned} D(a_1, \ldots, a_n) &= D(\lambda_2 a_2 + \ldots + \lambda_n a_n, a_2, \ldots, a_n) \\ &= \lambda_2 D(a_2, a_2, \ldots, a_n) + \ldots + \lambda_n D(a_n, a_2, \ldots, a_n). \end{aligned}$$

Pela propriedade (*i*), todos os termos na última linha são nulos.

$\square$

4.2.1 Determinantes e permutações

Nessa subseção mostraremos a fórmula clássica do determinante em termos de permutações. Para isso, consideremos vetores $a_1, \ldots, a_n \in \mathbb{R}^n$ arbitrários. Escrevendo cada um desses vetores em termos da base canônica do $\mathbb{R}^n$, obtemos

$$\begin{array}{rcl} a_1 & = & a_{11}e_1 + \ldots + a_{n1}e_n, \\ a_2 & = & a_{12}e_1 + \ldots + a_{n2}e_n, \\ \vdots & = & \vdots \\ a_n & = & a_{1n}e_1 + \ldots + a_{nn}e_n \end{array}$$

(estamos usando essa notação para os índices, pois os vetores $a_1, \ldots, a_n$ são colunas!).

Assim,

$$\begin{array}{rcl} D(a_1, \ldots, a_n) & = & D(a_{11}e_1 + \ldots + a_{n1}e_n, a_2, \ldots, a_n) \\ & = & a_{11}D(e_1, a_2, \ldots, a_n) + \ldots + a_{n1}D(e_n, a_2, \ldots, a_n). \end{array}$$

Se substituirmos agora a_2 por $a_{12}e_1 + \ldots + a_{n2}e_n$, obteremos uma expressão semelhante. Note que, a partir de cada parcela da soma acima, essa substituição cria outras n parcelas. Entretanto, existem alguns cancelamentos; por exemplo, sabemos que $D(e_1, e_1, a_3, \ldots, a_n) = 0$. Feitas todas as substituições de $a_2, \ldots, a_n$, chegaremos a

$$D(a_1, \ldots, a_n) = \sum_{i_1, \ldots, i_n = 1}^{n} a_{i_1 1} a_{i_2 2} \cdots a_{i_n n} D(e_{i_1}, \ldots, e_{i_n}).$$

Nesse somatório são nulas todas as parcelas em que há repetição de algum dos índices $i_1, \ldots, i_n$. De fato, nesse caso, temos que $i_k = i_j$ para $k \neq j$ e então $e_{i_k} = e_{i_j}$. A propriedade (i) do determinante garante então que $D(e_{i_1}, \ldots, e_{i_j}, \ldots, e_{i_k}, \ldots, e_{i_n}) = 0$. Quer dizer, como todos os índices $i_1, \ldots, i_n$ são diferentes entre si, está assim estabelecida uma permutação dos inteiros $\{1, \ldots, n\}$. Quer dizer, no somatório acima precisamos apenas considerar

$$D(a_1, \ldots, a_n) = \sum_p a_{p_1 1} a_{p_2 2} \cdots a_{p_n n} D(e_{p_1}, \ldots, e_{p_n}),$$

em que p percorre as permutações de $S = \{1, \ldots, n\}$. Se p é uma permutação, podemos escrevê-la como produto de transposições. Pela propriedade (iv) do determinante, uma transposição altera o valor de D pelo fator -1. Se p é um produto de k transposições, D será alterado por $(-1)^k$. Decorre então da fórmula (1) que $D(e_{p_1}, \ldots, e_{p_n}) = \sigma(p)D(e_1, \ldots, e_n) = \sigma(p)$, em virtude da propriedade (iii) do determinante. Assim temos, finalmente,

$$D(a_1, \ldots, a_n) = \sum_p \sigma(p) a_{p_1 1} a_{p_2 2} \cdots a_{p_n n}, \tag{4.2}$$

que é a expressão clássica do determinante em termos de permutações. (Muitos autores usam essa expressão como definição do determinante).

Exemplo 4.2.3 Sejam $a_1 = (a_{11}, a_{21})$ e $a_2 = (a_{12}, a_{22})$ vetores do $\mathbb{R}^2$. Calcule o determinante $D(a_1, a_2)$.

Precisamos, em primeiro lugar, determinar todas as permutações do conjunto $\{1, 2\}$. Elas são $p_1 = id$ e $p_2 = (1\ 2)$. Temos que $\sigma(p_1) = 1$ e $\sigma(p_2) = -1$ (note que p_2 é uma transposição!). Então

$$D(a_1, a_2) = \sum_p \sigma(p) a_{p_1 1} a_{p_2 2} = (1)a_{11}a_{22} + (-1)a_{12}a_{21} = a_{11}a_{22} - a_{12}a_{21}.$$

◀

O exercício 4 pede para que se calcule, dessa maneira, o determinante de três vetores genéricos do $\mathbb{R}^3$. É claro que, depois de ter feito esse exercício, você terá chegado à conclusão que esse processo para se calcular o determinante não é muito prático...Entretanto, o seguinte corolário é importante:

Corolário 4.2.4 *Existe (no máximo) uma função D satisfazendo às propriedades $(i) - (iii)$.*

Demonstração: De fato, se $\bar{D}$ também satisfizesse essas propriedades, $\bar{D}$ satisfaria a mesma expressão obtida para D em termos de permutações.

□

Falta, entretanto, mostrar que existe alguma função satisfazendo às propriedades do determinante. É o que faremos agora.

Teorema 4.2.5 *Existe uma única função determinante.*

Demonstração: Como vimos, basta provar a existência de uma função determinante. Mostraremos que a função definida pela expressão (4.2), isto é,

$$D(a_1, \ldots, a_n) = \sum_p \sigma(p) a_{p_1 1} a_{p_2 2} \cdots a_{p_n n}, \tag{4.3}$$

satisfaz às propriedades $(i) - (iii)$ da função determinante. (Lembramos que p denota uma permutação do conjunto $S = \{1, \ldots, n\}$).

(i) Suponhamos que os vetores a_i e a_j sejam iguais. Seja τ a transposição entre a_i e a_j. Então $a_{\tau p_1 1} a_{\tau p_2 2} \cdots a_{\tau p_n n} = a_{p_1 1} a_{p_2 2} \cdots a_{p_n n}$, pois τ transpõe os vetores a_i e a_j, que são iguais, e mantém fixos os outros vetores. Assim,

$$\begin{aligned} D(\ldots, a_i, \ldots, a_j, \ldots) &= \sum_p \sigma(p) a_{p_1 1} a_{p_2 2} \cdots a_{p_n n} = \sum_p \sigma(p) a_{\tau p_1 1} a_{\tau p_2 2} \cdots a_{\tau p_n n} \\ &= -\sum_p \sigma(\tau p) a_{\tau p_1 1} a_{\tau p_2 2} \cdots a_{\tau p_n n} = -D(\ldots, a_j, \ldots, a_i, \ldots), \end{aligned}$$

Como $D(\ldots, a_i, \ldots, a_j, \ldots) = D(\ldots, a_j, \ldots, a_i, \ldots)$, resulta que $D(\ldots, a_i, \ldots, a_j, \ldots) = 0$.

(ii) A linearidade é imediata, notando que cada parcela de (4.3) contém exatamente uma coordenada do vetor $a_i + kb_i$, de modo que

$$D(a_1, \ldots, a_i + kb_i, \ldots, a_n) = D(a_1, \ldots, a_i, \ldots, a_n) + kD(a_1, \ldots, b_i, \ldots, a_n).$$

(iii) Se $p_i \neq i$, a coordenada do vetor $a_i = e_i$ tomada no somatório (4.3) será nula. Assim, apenas a permutação identidade, produz termo não-nulo. No caso da identidade, temos sinal igual a 1 e todos os termos $a_{p_i i} = 1$, de modo que $D(e_1, \ldots, e_n) = 1$.

□

4.3 Propriedades do determinante de uma matriz

Definição 4.3.1 *Seja A uma matriz $n\times n$, com colunas $a_1,\ldots,a_n$. Definimos o* **determinante** *da matriz A por*

$$\det A = D(a_1,\ldots,a_n).$$

Nessa seção mostraremos algumas propriedades clássicas do determinante de uma matriz.

4.3.1 O determinante da matriz transposta

Uma vez que $\sigma(id)=1$, notamos que $1=\sigma(pp^{-1})=\sigma(p)\sigma(p^{-1})$, o que mostra que $\sigma(p)=\sigma(p^{-1})$.

Teorema 4.3.2 *Seja A uma matriz $n\times n$ e A^T a transposta da matriz A. Então*

$$\det A = \det A^T.$$

Demonstração: A equação (4.3) garante que

$$\det A = \sum_p \sigma(p) a_{p_1 1} a_{p_2 2}\cdots a_{p_n n}.$$

Mas, se $p(i)=j$, então $i=p^{-1}p(i)=p^{-1}(j)$. Como estamos denotando $p(i)=p_i$, denotaremos $p^{-1}(j)={p^{-1}}_j$, de modo que a última expressão pode ser escrita como, $i={p^{-1}}_j$. Assim, se $p_1=j$, então $a_{p_1 1}=a_{j{p^{-1}}_j}$. Da mesma forma para os outros índices, de modo que

$$\sum_p \sigma(p) a_{p_1 1} a_{p_2 2}\cdots a_{p_n n} = \sum_p \sigma(p) a_{1{p^{-1}}_1} a_{2{p^{-1}}_2}\cdots a_{n{p^{-1}}_n}.$$

Mas se p percorre todas as permutações de $\{1,\ldots,n\}$, o mesmo acontece com p^{-1}. Uma vez que o sinal de p e o de p^{-1} é o mesmo, chegamos a

$$\det A = \sum_{p^{-1}} \sigma(p^{-1}) a_{1{p^{-1}}_1} a_{2{p^{-1}}_2}\cdots a_{n{p^{-1}}_n} = \sum_p \sigma(p) a_{1p_1} a_{2p_2}\cdots a_{np_n},$$

que é o determinante da matriz transposta, pois cada uma de suas entradas aparece na forma a_{ji} ao invés de a_{ij}.

$\square$

4.3.2 O determinante do produto de matrizes quadradas

Teorema 4.3.3 *Sejam A,B matrizes $n\times n$. Então*

$$\det(BA) = \det A \det B.$$

Demonstração: A equação (3.3) garante que a j-ésima coluna da matriz BA é $(BA)e_j$. Da mesma forma, a j-ésima coluna de A é $Ae_j = a_j$. Assim, a j-ésima coluna de BA é

$$(BA)e_j = B(Ae_j) = Ba_j.$$

Temos então, por definição,

$$\det(BA) = D(Ba_1, \ldots, Ba_n).$$

Suponhamos que $\det B \neq 0$. Definimos então a função C por

$$C(a_1, \ldots, a_n) = \frac{\det(BA)}{\det B}.$$

Em virtude da expressão para $\det(BA)$ obtida acima, podemos escrever C como

$$C(a_1, \ldots, a_n) = \frac{D(Ba_1, \ldots, Ba_n)}{\det B}. \tag{4.4}$$

Vamos provar que a função C satisfaz às propriedades $(i) - (iii)$ postuladas para a função determinante. Temos

(i) Se $a_i = a_j$, para $i \neq j$, então $Ba_i = Ba_j$. Como D satisfaz à propriedade (i), temos $C = 0$;

(ii) Como $B(x + \lambda y) = Bx + \lambda By$, cada Ba_i é uma função linear de a_i. Como D é n-linear, o mesmo vale para C;

(iii) Para $a_i = e_i$, temos

$$C(e_1, \ldots, e_n) = \frac{D(Be_1, \ldots, Be_n)}{\det B}.$$

Mas $Be_i = b_i$, a i-ésima coluna de B. Logo

$$C(e_1, \ldots, e_n) = \frac{D(b_1, \ldots, b_n)}{\det B} = \frac{\det B}{\det B} = 1.$$

Uma vez que existe uma única função determinante, $C(a_1, \ldots, a_n) = D(a_1, \ldots, a_n)$. Mas isso prova o afirmado, quando $\det B \neq 0$.

Quando $\det B = 0$ definimos, para cada t, a matriz $B(t) = B + tI$. Claramente $\det(B(0)) = \det B = 0$. A expressão (4.2) mostra que $D(B(t))$ é um polinômio de grau n e que o coeficiente do termo t^n é igual a 1 (vide exercício 6). Isso quer dizer que $D(B(t)) = 0$ para no máximo n valores de t. Em particular, $D(B(t)) \neq 0$ para todos os valores de t suficientemente próximos de 0, mas diferentes de zero. Assim, pelo que já mostramos,

$$\det(B(t)A) = \det A \det(B(t)).$$

Fazendo t tender a zero, chegamos a $\det(BA) = 0$, como queríamos[2].

□

[2] Veja exercício 6.

4.3.3 O determinante em termos de cofatores

Lema 4.3.4 *Seja A uma matriz cuja primeira coluna é o vetor e_1:*

$$A = \begin{pmatrix} 1 & * & \cdots & * \\ 0 & & & \\ \vdots & & A_{11} & \\ 0 & & & \end{pmatrix},$$

em que A_{11} denota a submatriz $(n-1)\times(n-1)$ formada pelos a_{ij} com $i,j>1$. Então vale

$$\det A = \det A_{11}.$$

Demonstração: Notamos inicialmente que podemos considerar apenas o caso em que

$$A = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & A_{11} & \\ 0 & & & \end{pmatrix}.$$

De fato, decorre das propriedades (i) e (ii) que o determinante de A não é alterado se somarmos a cada uma das colunas $a_2,\ldots,a_n$ um múltiplo adequado a primeira coluna. Ora, com esse processo conseguimos zerar a primeira linha da matriz A.

Para a matriz A com zeros na primeira coordenada de todas as colunas exceto da primeira, definimos, como função das colunas da matriz A_{11},

$$C(A_{11}) = \det \begin{pmatrix} 1 & 0 \\ 0 & A_{11} \end{pmatrix}.$$

Essa função satisfaz às propriedades $(i)-(iii)$ da definição do determinante. A unicidade do determinante então garante que $\det C(A_{11}) = \det A_{11}$. Mas isso prova o afirmado. (Note que estamos repetindo o argumento utilizado na equação (4.4)).

□

Corolário 4.3.5 *Seja A uma matriz cuja j-ésima coluna é o vetor e_i. Então*

$$\det A = (-1)^{i+j} \det A_{ij},$$

em que A_{ij} é a matriz obtida de A ao se eliminar sua i-ésima linha e sua j-ésima coluna. Em outras palavras, A_{ij} é o menor (ij) de A.

Demonstração: Como cada troca de uma coluna por outra corresponde a uma multiplicação do determinante por (-1), após j trocas, faremos que o vetor e_i esteja na primeira coluna da matriz A e as colunas restantes mantendo sua ordem original. Quer dizer,

$$\det A = (-1)^j D(e_i, a_1, \cdots, a_{j-1}, a_{j+1}, \cdots, a_n).$$

Como o determinante de A e de A^T são iguais (pelo teorema 4.3.2), também podemos alterar a ordem das linhas da matriz A. Após i trocas (o que resulta em multiplicação do determinante por $(-1)^i$), o vetor e_i terá se transformado no vetor e_1 e a matriz resultante estará como no lema 4.3.4. Daí decorre o afirmado.

□

Agora mostraremos a expansão clássica do determinante em termos de uma de suas colunas.

Teorema 4.3.6 *Seja A uma matriz $n \times n$ e $j \in \{1, \ldots, n\}$. Então*

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij}.$$

Demonstração: Para simplificar a notação, vamos considerar $j = 1$. Escrevendo a_1 como combinação linear dos vetores da base canônica, temos

$$a_1 = a_{11}e_1 + \cdots + a_{n1}e_n.$$

Como D é n-linear, temos

$$\det A = D(a_1, \ldots, a_n) = a_{11}D(e_1, a_2, \ldots, a_n) + \cdots + a_{n1}D(e_n, a_2, \ldots, a_n).$$

O resultado segue então do corolário 4.3.5.

□

4.4 A regra de Cramer

Consideremos a expressão

$$Ax = u,$$

em que A é uma matriz $n \times n$. Podemos interpretá-la, supondo conhecido o vetor x, como a definição de u. Por outro lado, conhecido u, ela pode ser vista como uma equação na variável x.

Suponhamos conhecido o vetor x. Exprimindo x em termos dos vetores da base canônica, $x = \sum_{j=1}^{n} x_j e_j$, a relação $Ae_j = a_j$ (em que a_j é a j-ésima coluna de A) garante que

$$\sum_{j=1}^{n} x_j a_j = u.$$

Definimos agora a matriz A_k, obtida ao se substituir a k-ésima coluna de A pelo vetor u. Então, descrevendo essa matriz em termos de suas colunas,

$$\begin{aligned} A_k &= (a_1 \; \ldots \; a_{k-1} \; u \; a_{k+1} \; \ldots \; a_n) \\ &= (a_1 \; \ldots \; a_{k-1} \; \sum_{j=1}^{n} x_j a_j \; a_{k+1} \; \ldots \; a_n). \end{aligned}$$

Assim, se D é a função determinante,

$$\det A_k = \sum_{j=1}^{n} x_j \det(a_1 \ldots, a_{k-1}, a_j, a_{k+1}, \ldots, a_n).$$

Logo,

$$\det A_k = x_k \det A,$$

pois apenas esse termo não se anula no somatório, pela propriedade (i) da definição da função determinante. Portanto,

$$x_k = \frac{\det A_k}{\det A}, \tag{4.5}$$

desde que $\det A \neq 0$. Essa é a regra de Cramer para se obter a solução da equação $Ax = u$, para um dado u. Ela garante que, se $\det A \neq 0$, então a (única) solução x de $Ax = u$ é dada por $x_k = \frac{\det A_k}{\det A}$.

Se expandirmos $\det A_k$ com relação a sua k-ésima coluna, obtemos

$$\det A_k = \sum_{i=1}^{n} (-1)^{i+k} \det A_{ik} u_i.$$

Dividindo essa equação por $\det A$, encontramos então

$$x_k = \frac{\det A_k}{\det A} = \sum_{i=1}^{n} (-1)^{i+k} \frac{\det A_{ik}}{\det A} u_i. \tag{4.6}$$

Vamos escrever a expressão acima em linguagem matricial.

Teorema 4.4.1 *A matriz A é invertível se, e somente se,* $\det A \neq 0$. *Nesse caso, a matriz inversa A^{-1} tem a forma*

$$(A^{-1})_{ki} = (-1)^{i+k} \frac{\det A_{ik}}{\det A}. \tag{4.7}$$

Demonstração: Suponhamos que A tenha inversa. Então existe A^{-1} com $AA^{-1} = I$. Aplicando o determinante em ambos os lados dessa igualdade, obtemos $\det A \det A^{-1} = \det I = 1$. Logo, $\det A \neq 0$.

Reciprocamente, suponhamos que $\det A \neq 0$. Então a equação (4.7) faz sentido, definindo assim uma matriz que, por abuso de linguagem, chamaremos A^{-1}. Façamos essa matriz agir sobre u. Pela fórmula (3.1), temos

$$(A^{-1}u)_k = \sum_{i=1}^{n} (A^{-1})_{ki} u_i.$$

Substituindo (4.7) nessa fórmula e comparando com (4.6), obtemos

$$(A^{-1}u)_k = x_k,$$

o que mostra que A^{-1} é realmente a inversa da matriz A.

□

4.5 Matrizes semelhantes

Definição 4.5.1 *Seja $A = (a_{ij})$ uma matriz quadrada. Definimos o* **traço** *da matriz A, denotado* $\operatorname{tr} A$*, por*

$$\operatorname{tr} A = \sum_{i=1}^{n} a_{ii}.$$

Teorema 4.5.2 *O traço é uma aplicação linear e* $\operatorname{tr}(AB) = \operatorname{tr}(BA)$*.*

Demonstração: A linearidade do traço é óbvia. Por definição, temos

$$(AB)_{ii} = \sum_{k=1}^{n} a_{ik}b_{ki} \quad \text{e} \quad (BA)_{kk} = \sum_{i=1}^{n} b_{ki}a_{ik}.$$

Assim,

$$\operatorname{tr}(AB) = \sum_{i=1}^{n}\left(\sum_{k=1}^{n} a_{ik}b_{ki}\right) = \sum_{k=1}^{n}\left(\sum_{i=1}^{n} b_{ki}a_{ik}\right) = \operatorname{tr}(BA).$$

□

Definição 4.5.3 *Duas matrizes A e B são* **semelhantes***, denotado $A \sim B$, se existe uma matriz invertível P tal que $B = P^{-1}AP$.*

Claramente temos assim definida uma relação de equivalência[3] no conjunto das matrizes $n \times n$.

Teorema 4.5.4 *Matrizes semelhantes possuem o mesmo determinante e o mesmo traço.*

Demonstração: Temos

$$\det B = \det(P^{-1}AP) = \det P^{-1} \det A \det P = \det A \det(P^{-1}P) = \det A \det I = \det A.$$

Também, pelo teorema 4.5.2,

$$\operatorname{tr} B = \operatorname{tr}(P^{-1}AP) = \operatorname{tr}(APP^{-1}) = \operatorname{tr}(AI) = \operatorname{tr} A.$$

□

Como vimos anteriormente, dada uma aplicação linear T de um espaço X de dimensão n nele mesmo, ao se escolher uma base de X, podemos representar T por uma matriz. Duas representações de T, obtidas pela escolha de duas bases distintas, são semelhantes. Aplicando o teorema anterior, vemos que faz sentido a seguinte definição:

Definição 4.5.5 *Seja $T : V \to V$ uma aplicação linear definida no espaço vetorial de dimensão finita V. Definimos*

$$\operatorname{tr} T = \operatorname{tr} T_{\mathcal{B}}^{\mathcal{B}} = \operatorname{tr} T_{\mathcal{B}} \quad e \quad \det T = \det T_{\mathcal{B}}^{\mathcal{B}} = \det T_{\mathcal{B}},$$

em que $\mathcal{B}$ é qualquer base de V.

---

[3]Veja exercício 21 do capítulo 3.

4.6 Exercícios

1. Seja $p : S \to S$ uma permutação e $\{a, p(a), \ldots, p^{k-1}(a)\}$ os elementos da órbita de a. Se b é um elemento da órbita de a, mostre que $p^k(b) = b$. Em outras palavras, todos os elementos da órbita de a têm a mesma ordem.

2. Mostre que o sinal de uma transposição é igual a -1.

3. Mostre que a propriedade (iv) da função determinante implica a propriedade (i). Assim, poderíamos ter definido o determinante como uma função que satisfaz às propriedades $(ii) - (iii) - (iv)$.

4. Repita o exemplo 4.2.3 para três vetores genéricos do $\mathbb{R}^3$. Em outras palavras, calcule o determinante de uma matriz 3×3.

5. Aplique as propriedades da função determinante para calcular o determinante da matriz
$$\begin{pmatrix} 2 & 5 & -3 & -2 \\ -2 & -3 & 2 & -5 \\ 1 & 3 & -2 & 2 \\ -1 & -6 & 4 & 3 \end{pmatrix}.$$

6. Se B é uma matriz $n \times n$ e $B(t) = B + tI$, mostre que $\det(B(t))$ é um polinômio mônico[4] de grau n na variável t e, portanto, uma função contínua. Se A é uma matriz $n \times n$, verifique que $\det(AB(t))$ também é um polinômio de grau n em t. Mostre que a função determinante é uma função contínua e justifique, assim, as propriedade de limite utilizadas na demonstração do teorema 4.3.3.

7. Seja A uma matriz **triangular superior**, isto é, uma matriz da forma
$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}.$$
Mostre que $\det A = a_{11} \cdots a_{nn}$. A transposta da matriz A é uma matriz **triangular inferior**.

8. Seja
$$A = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_j \end{pmatrix},$$

[4]isto é, o coeficiente do termo de maior grau é igual a 1.

em que cada A_i $(i = 1, \ldots, j)$ é uma matriz quadrada. Mostre que $\det A = \det A_1 \cdots \det A_j$. Generalize para uma matriz A que seja triangular superior por blocos, isso é, uma matriz da forma

$$A = \begin{pmatrix} A_1 & * & \cdots & * \\ 0 & A_2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_j \end{pmatrix},$$

em que $*$ denota uma matriz de ordem adequada.

9. Resolva, utilizando a regra de Cramer, o sistema

$$\begin{array}{rcrcrcrcr} 2x_1 & + & x_2 & + & 5x_3 & + & x_4 & = & 5 \\ x_1 & + & x_2 & - & 3x_3 & - & 4x_4 & = & -1 \\ 3x_1 & + & 6x_2 & - & 2x_3 & + & x_4 & = & 8 \\ 2x_1 & + & 2x_2 & + & 2x_3 & - & 3x_4 & = & 2 \end{array}$$

10. Usando a regra de Cramer, determine os valores de k para os quais o sistema

$$\begin{array}{rcrcrcr} kx & + & y & + & z & = & 1 \\ x & + & ky & + & z & = & 1 \\ x & + & y & + & kz & = & 1 \end{array}$$

possui solução única. Compare com o resultado obtido através de escalonamento (método de Gauss - veja a seção 8.1).

11. Utilizando escalonamento, dê uma demonstração alternativa para o teorema 4.4.1.

Capítulo 5

Teoria Espectral

5.1 Autovetores e autovalores

Dados um espaço vetorial V de dimensão finita sobre o corpo $\mathbb{K}$ e uma aplicação linear $T : V \to V$, queremos encontrar uma base $\mathcal{B}$ de V tal que a representação $T_{\mathcal{B}}$ desse operador na base $\mathcal{B}$ seja a mais simples possível.

Consideremos a seguinte situação: suponhamos que se tenha

$$V = W_1 \oplus W_2 \oplus \cdots \oplus W_n, \quad \dim W_i = 1, \quad TW_i \subset W_i, \quad 1 \leq i \leq n. \tag{5.1}$$

Seja $\{w_i\}$ uma base de W_i. Então $\mathcal{B} = \{w_1, \ldots, w_n\}$ é uma base de V. Como $T(W_i) \subset W_i$, existe $\lambda_i \in \mathbb{K}$ tal que $Tw_i = \lambda_i w_i$ (não estamos supondo que $\lambda_i \neq \lambda_j$ para $i \neq j$). A representação de T na base $\mathcal{B}$ (no domínio e na imagem) é a matriz $A = T_{\mathcal{B}}$

$$A = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

Dizemos então que T é **diagonalizável**. Note que, se T for diagonalizável, então vale a decomposição (5.1), resultado que formalizaremos no teorema 5.1.4, mais abaixo.

Observe que a igualdade $Tw_i = \lambda_i w_i$ garante que $(\lambda_i I - T)$ não é um operador injetivo; portanto, $(\lambda_i I - T)$ não é um isomorfismo. Assim, $\det(\lambda_i I - T) = \det(\lambda_i I - A) = 0$, de acordo com a definição 4.5.5. Isso quer dizer que λ_i é uma raiz do polinômio (na variável t) $p(t) = \det(tI - A)$, chamado **polinômio característico** do operador T (ou da matriz A). Lembramos[1] que $p(t)$ é um polinômio mônico de grau n. Assim, como esse polinômio possui n raízes no corpo $\mathbb{K}$, podemos concluir (mesmo quando $\lambda_i = \lambda_j$ para $i \neq j$) que

$$p(t) = (t - \lambda_1)(t - \lambda_2) \cdots (t - \lambda_n) \tag{5.2}$$

e $w_i \in \ker(T - \lambda_i I)$. Note que a equação $Tx = \lambda_i x$ é satisfeita para qualquer elemento de W_i.

[1] Veja exercício 6 do capítulo 4.

Consideremos o operador T e seu polinômio característico $p(t) = \det(tI - T)$. As raízes $\lambda \in \mathbb{K}$ do polinômio característico são chamadas **autovalores**[2] de T. Se existirem n raízes distintas, isto é, se

$$p(t) = (t - \lambda_1) \cdots (t - \lambda_n),$$

com $\lambda_i \neq \lambda_j$ para $i \neq j$, o espaço $W_i := \ker(T - \lambda_i I)$ terá dimensão 1. De fato, existe pelo menos um vetor não-nulo w_i tal que $(\lambda_i I - T)w_i = 0$ pois, como $\lambda_i I - T$ não tem inversa, o sistema $(\lambda_i I - T)x = 0$ tem solução não-trivial w_i. O vetor $0 \neq w_i \in W_i$, solução de $(\lambda_i I - T)x = 0$, é chamado **autovetor** de T associado ao autovalor λ_i. Agora afirmamos: autovetores w_i associados a autovalores distintos são linearmente independentes. (Aceitaremos isso momentaneamente). Mas então $\{w_1, \ldots, w_n\}$ é uma base de V. Quer dizer, nesse caso especial em que o polinômio característico possui n raízes distintas, teremos provado que

$$V = W_1 \oplus W_2 \oplus \cdots \oplus W_n,$$

com $W_i = \ker(T - \lambda_i I)$ e $T(W_i) \subset W_i$ (pois $T(cw_i) = cTw_i = c\lambda w_i \in W_i$).

Estamos agora na situação em que iniciamos e, portanto, a representação de T na base $\mathcal{B} = \{w_1, \ldots, w_n\}$ será justamente a matriz diagonal dada por A.

Entretanto, nem sempre o polinômio característico é produto de fatores lineares distintos, mesmo quando o operador T é diagonalizável. Considere o seguinte exemplo:

Exemplo 5.1.1 Para a aplicação identidade $I : \mathbb{K}^n \to \mathbb{K}^n$, o polinômio característico $p(\lambda) = \det(I - \lambda I) = (1 - \lambda)^n$. Quer dizer, $\lambda_1 = \cdots = \lambda_n$. Temos a decomposição (5.1)

$$\mathbb{K}^n = W_1 \oplus \cdots \oplus W_n$$

com $W_i = \{ce_i;\, c \in \mathbb{K}\}$, mas $\ker(I - 1I) = \ker 0 = \mathbb{K}^n$. ◀

Antes de mostrarmos que autovetores associados a autovalores distintos são linearmente independentes, daremos algumas definições e estudaremos uma situação simples.

Definição 5.1.2 *Sejam V um espaço vetorial sobre o corpo $\mathbb{K}$, com dimensão finita n, e $T : V \to V$ uma aplicação linear. O polinômio $p(t) := \det(tI - T)$ é o* **polinômio característico** *de T. As raízes $\lambda_i \in \mathbb{K}$ desse polinômio são chamadas* **autovalores** *de T. Os elementos não nulos do núcleo $\ker(T - \lambda_i I)$ são chamados* **autovetores** *associados ao autovalor λ_i, ou simplesmente autovetores de T. O núcleo*

$$\ker(T - \lambda_i I) = \{v \in V;\, (T - \lambda_i I)v = 0\}$$

é o **autoespaço** *associado ao autovalor λ_i.*

[2]O nome **espectro**, usual no estudo de operadores lineares em espaços de dimensão infinita, refere-se a uma generalização do conceito de autovalor: dado um operador linear $T : V \to V$, o espectro de T é formado por escalares λ tais que: *(i)* $(T - \lambda I)$ não é injetivo; *(ii)* $(T - \lambda I)$ não é sobrejetivo; *(iii)* $(T - \lambda I)^{-1}$ não é contínuo. As duas primeiras opções são equivalentes em espaços de dimensão finita; a terceira nunca ocorre nestes espaços. Assim, o espectro de um operador é o mesmo que o conjunto de autovalores de um operador linear, no caso de dimensão finita.

Frisamos que apenas as raízes $\lambda_i \in \mathbb{K}$ do polinômio característico são autovalores do operador. Assim, se $T : \mathbb{R}^2 \to \mathbb{R}^2$ é definido por $T(x, y) = (-y, x)$, então seu polinômio característico é $p(t) = t^2 + 1$, que não possui raízes reais. Portanto, T não possui autovalores. Considerando $T : \mathbb{C}^2 \to \mathbb{C}^2$ definido da mesma maneira, $p(t) = t^2 + 1 = (t-i)(t+i)$, e T possui dois autovalores distintos. Isso mostra que a análise de uma aplicação linear $T : V \to V$ depende muito do corpo $\mathbb{K}$ sobre o qual V é espaço vetorial.

Observação 5.1.3 O polinômio característico de $T : V \to V$ é especialmente importante por causa de suas raízes, os autovalores de T. Como $\det(T - tI) = (-1)^n \det(tI - T)$ (em que n é a dimensão de V) possui as mesmas raízes, também é usual chamar de polinômio característico de T ao polinômio $\det(T - tI)$. ◀

Teorema 5.1.4 *Uma aplicação linear $T : V \to V$ pode ser representada por uma matriz diagonal se, e somente se, existe uma base $\mathcal{B}$ de V formada por autovetores de T.*

Demonstração: Suponhamos que $\mathcal{B} = \{v_1, \ldots, v_n\}$ seja uma base de V tal que $T_{\mathcal{B}}^{\mathcal{B}} = T_{\mathcal{B}}$ seja uma matriz diagonal:

$$T_{\mathcal{B}} = D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

(não estamos supondo que os λ_i sejam distintos!).

Sabemos que a i-ésima coluna de D é $[T(v_i)]_{\mathcal{B}} = \lambda_i e_i$. Como e_i é a representação de v_i na base $\mathcal{B}$, mostramos que $T(v_i) = \lambda_i v_i$.

A recíproca é imediata.

□

(Veja, a esse respeito, o exercício 3).

Mostraremos agora o fato utilizado anteriormente.

Teorema 5.1.5 *Autovetores de T correspondentes a autovalores distintos são linearmente independentes.*

Demonstração: Sejam w_i, $1 \leq i \leq n$, autovetores de T associados a autovalores distintos λ_i. Faremos indução no número n. Se $n = 1$, o resultado é óbvio. Suponhamos verdadeiro para $n = j - 1$ e consideremos $n = j$. Se

$$\alpha_1 w_1 + \alpha_2 w_2 + \ldots + \alpha_j w_j = 0, \tag{5.3}$$

aplicando T em (5.3), obtemos

$$\alpha_1 T w_1 + \alpha_2 T w_2 + \ldots + \alpha_j T w_j = 0.$$

Mas $T w_i = \lambda_i w_i$. Assim,

$$\alpha_1 \lambda_1 w_1 + \ldots + \alpha_n \lambda_j w_j = 0.$$

Por outro lado, multiplicando (5.3) por λ_j, obtemos

$$\alpha_1\lambda_j w_1 + \alpha_2\lambda_j w_2 + \ldots + \alpha_j\lambda_j w_j = 0.$$

Subtraindo essas duas últimas equações, vem

$$\alpha_1(\lambda_1 - \lambda_j)w_1 + \alpha_2(\lambda_2 - \lambda_j)w_2 + \ldots + \alpha_{j-1}(\lambda_{j-1} - \lambda_j)w_{j-1} = 0.$$

Como $\lambda_i - \lambda_j \neq 0$ para todo $i = 1, \ldots, j-1$, a hipótese de indução garante que $\alpha_i = 0$ para esses valores de i. Levando em (5.3), concluímos que $\alpha_j = 0$ e que os vetores são linearmente independentes.

□

Na verdade, temos a seguinte generalização do teorema anterior:

Corolário 5.1.6 *Se $\lambda_1, \ldots, \lambda_j$ são autovalores distintos de T e $W_i = \ker(T - \lambda_i I)$, então o subespaço $W = W_1 + \cdots + W_j$ é a soma direta dos subespaços W_i, ou seja,*

$$W = W_1 \oplus \cdots \oplus W_j.$$

Em outras palavras, sejam $w_{i1}, \ldots, w_{ik_i}$ autovetores linearmente independentes associados ao autovalor λ_i da aplicação linear T, com $i = 1, \ldots, j$. Então o conjunto

$$\{w_{11}, w_{12}, \ldots, w_{1k_1}, w_{21}, \ldots, w_{2k_2}, \ldots, w_{j1}, \ldots, w_{jk_j}\}$$

é linearmente independente.

Demonstração: Escolha $w_i \in W_i$, para $i = 1, \ldots, j$. Cada $w_i \neq 0$ é um autovetor associado a λ_i. Pelo teorema 5.1.5 temos que $\alpha_1 w_1 + \ldots + \alpha_j w_j = 0$ se, e somente se, cada parcela $\alpha_i w_i = 0$, para todo i. Isso é o mesmo que afirmar a unicidade da decomposição $w = w_1 + \ldots + w_j$.

□

O corolário acima é importante, pois podemos ter vários autovetores linearmente independentes associados ao mesmo autovalor. Essa é a situação que ocorre no exemplo 5.1.1.

Para finalizar essa seção, enunciamos o resultado que mostramos anteriormente:

Corolário 5.1.7 *Se V é um espaço vetorial de dimensão n e o polinômio característico da aplicação linear $T : V \to V$ possui n raízes distintas, então V possui uma base $\mathcal{B}$ formada por autovetores de T. A aplicação T representada na base $\mathcal{B}$ é uma matriz diagonal, sendo os elementos da diagonal principal os autovalores de T.*

5.2 Polinômios de aplicações lineares

Sejam V um espaço vetorial sobre o corpo $\mathbb{K}$, $T : V \to V$ uma aplicação linear e $q(t)$ um polinômio com coeficientes no corpo $\mathbb{K}$ (denotaremos então $q \in \mathbb{K}[t]$). Se $q(t) = a_k t^k + a_{k-1}t^{k-1} + a_1 t + a_0$, claramente faz sentido calcular

$$q(T) := a_k T^k + a_{k-1}T^{k-1} + \cdots + a_1 T + a_0 I,$$

mesmo que V tenha dimensão infinita. Se $A \in \mathbb{M}_{n\times n}$ é uma matriz, $q(A)$ é uma matriz $n \times n$. Vamos mostrar como os autovalores de A se relacionam com os autovalores de $q(A)$.

Teorema 5.2.1 (Teorema da Imagem do Espectro)[3]

Sejam $A \in \mathbb{M}_{n\times n}(\mathbb{K})$ e $q \in \mathbb{K}[t]$. Se λ é um autovalor de A, então $q(\lambda)$ é um autovalor de $q(A)$. Além disso, quando $\mathbb{K} = \mathbb{C}$, todos os autovalores de $q(A)$ são da forma $q(\lambda)$, em que λ é um autovalor de A.

Demonstração: Se λ é um autovalor de A, então existe $v \neq 0$ tal que $Av = \lambda v$. Decorre daí que $A^j v = \lambda^j v$ para todo $j \in \mathbb{N}$. Se $q(A) = a_k A^k + a_{k-1}A^{k-1} + \cdots + a_1 A + a_0 I$, então

$$q(A)v = a_k A^k v + \ldots + a_1 Av + a_0 Iv = a_k \lambda^k v + \ldots + a_1 \lambda v + a_0 v = q(\lambda)v.$$

Isso é o mesmo que afirmar que $(q(A) - q(\lambda)I)v = 0$, ou seja, que v é um autovetor de $q(A)$ associado ao autovalor $q(\lambda)$.

Reciprocamente, suponhamos que $\mathbb{K} = \mathbb{C}$ e que μ seja um autovalor de $q(A)$. Queremos mostrar que $\mu = q(\lambda)$, para algum autovalor λ da matriz A. Consideremos o polinômio $q(t) - \mu$. Fatorando esse polinômio no corpo $\mathbb{C}$, temos

$$q(t) - \mu = a_k \prod_{i=1}^{k} (t - r_i),$$

em que r_i são as raízes (não necessariamente distintas) desse polinômio (note que a_k é o coeficiente do termo de maior grau de $q(t)$). Temos então que (note que essa expressão independe de μ ser autovalor de $q(A)$)

$$q(A) - \mu I = a_k \prod_{i=1}^{k} (A - r_i I). \tag{5.4}$$

Como μ é um autovalor de $q(A)$, a matriz do lado esquerdo da igualdade não tem inversa. Isso quer dizer que ao menos uma das matrizes $(A - r_i I)$ não tem inversa e, portanto, r_i é um autovalor de A. Como r_i é raiz de $q(t) - \mu$, temos $q(r_i) - \mu = 0$, ou seja, $q(r_i) = \mu$.

□

Dizemos que dois polinômios $p,\ q \in \mathbb{K}[t]$ são **primos entre si** se o único polinômio mônico que divide tanto p quanto q é o polinômio 1.

Observação 5.2.2 Seja A uma matriz real. Como $\mathbb{R} \subset \mathbb{C}$, podemos vê-la como uma matriz sobre $\mathbb{C}$. Se $\lambda \in \mathbb{C} \setminus \mathbb{R}$ é uma raiz do polinômio característico $p(t) = a_0 + a_1 t + \ldots + a_n t^n$, então $p(\bar{\lambda}) = 0$. De fato, como todos os coeficientes desse polinômio são reais, o resultado segue quando tomamos os conjugados na equação $p(\lambda) = 0$. Como conseqüência, fatores quadráticos que aparecem na decomposição do polinômio característico em fatores irredutíveis surgem do produto $(t - \lambda)(t - \bar{\lambda})$ das raízes $\lambda, \bar{\lambda} \in \mathbb{C} \setminus \mathbb{R}$. ◀

Lema 5.2.3 *Sejam $p, q \in \mathbb{K}[t]$. Se p e q são primos entre si, então existem polinômios $a, b \in \mathbb{K}[t]$ tais que*

$$ap + bq = 1.$$

[3]Em toda a bibliografia consultada, nunca encontrei uma tradução para "Spectral Mapping Theorem". Acho inadequada a tradução "Teorema da Aplicação Espectral".

Demonstração: Seja $\mathcal{I}$ o conjunto de todos os polinômios da forma $ap + bq$, com $a, b \in \mathbb{K}[t]$. Como $\mathcal{I}$ possui elemento não nulo, existe em $\mathcal{I}$ um polinômio não nulo de menor grau, que chamaremos $d = \overline{a}p + \overline{b}q$.

Afirmamos que d divide tanto p quanto q. De fato, se d não dividisse p, por exemplo, teríamos $p = md + r$, em que o grau de r é menor do que o grau de d. Como p e d estão em $\mathcal{I}$, $r = p - md \in \mathcal{I}$, o que contradiz a escolha de d. Logo $r = 0$, mostrando o afirmado.

Como p e q são primos entre si, d tem grau zero, isto é, d é uma constante, digamos k. Como $d \neq 0$, escolhendo $a = \overline{a}/k$ e $b = \overline{b}/k$ temos

$$ap + bq = 1.$$

□

Corolário 5.2.4 *Se $p_1, \ldots, p_k, p_{k+1} \in \mathbb{K}[t]$ são primos entre si dois a dois, então $p_2 \ldots p_k p_{k+1}$ e p_1 são primos entre si.*

Demonstração: Isso se prova por indução em k. Se $k = 1$, nada há a provar. Suponhamos verdadeiro para $k = j$ e seja d um polinômio mônico que divide p_1 e $p_2 \ldots p_j p_{j+1}$. Como p_1 e p_{j+1} são primos entre si, existem polinômios a e b tais que $ap_1 + bp_{j+1} = 1$. Multiplicando por $p_2 \ldots p_j$, obtemos $ap_1(p_2 \ldots p_j) + b(p_2 \ldots p_j p_{j+1}) = p_2 \ldots p_j$. Como d divide tanto p_1 quanto $p_2 \ldots p_j p_{j+1}$, vemos que d divide $p_2 \ldots p_j$. Mas então a hipótese de indução garante que $d = 1$, provando o afirmado.

□

Lema 5.2.5 *Sejam $p, q \in \mathbb{K}[t]$ primos entre si e $0 \neq A \in \mathbb{M}_{n\times n}(\mathbb{K})$. Sejam N_p, N_q e N_{pq} os núcleos das matrizes $p(A)$, $q(A)$ e $p(A)q(A)$, respectivamente. Então*

$$N_{pq} = N_p \oplus N_q.$$

Demonstração: Como existem polinômios $a, b \in \mathbb{K}[t]$ tais que $bq + ap = 1$, temos que

$$b(A)q(A) + a(A)p(A) = I.$$

Se $x \in N_{pq}$, então $b(A)q(A)x \in N_p$. De fato, aplicando $p(A)$ a esse ponto, temos $p(A)b(A)q(A)x = b(A)p(A)q(A)x = 0$, dada a comutatividade de polinômios da matriz A. Da mesma forma temos $a(A)p(A)x \in N_q$, se $x \in N_{pq}$. Como $b(A)q(A)x + a(A)p(A)x = x$, mostramos que $x = x_p + x_q$, com $x_p \in N_p$ e $x_q \in N_q$.

Para mostrar que essa decomposição é única, suponhamos que $x = x_p + x_q = \overline{x}_p + \overline{x}_q$. Mas então $y := x_p - \overline{x}_p = \overline{x}_q - x_q$ pertence, simultaneamente, a N_p e N_q. Aplicando $b(A)q(A) + a(A)p(A) = I$ em y, temos

$$b(A)q(A)y + a(A)p(A)y = y.$$

Mas $b(A)q(A)y = 0 = a(A)p(A)y$, de modo que $y = 0$, o que implica $x = \overline{x}_p$ e $x_q = \overline{x}_q$, mostrando a unicidade da decomposição.

□

Por indução, obtemos então o

Corolário 5.2.6 *Seja $0 \neq A \in \mathbb{M}_{n\times n}(\mathbb{K})$. Se $p_1, p_2, \ldots, p_k$ são polinômios em $\mathbb{K}[t]$, primos entre si dois a dois, se N_{p_i} denota o núcleo de $p_i(A)$ e $N_{p_1\ldots p_k}$ o núcleo de $p_1(A)\ldots p_k(A)$, então*

$$N_{p_1\ldots p_k} = N_{p_1} \oplus \cdots \oplus N_{p_k}.$$

O **polinômio mínimo** $m \in \mathbb{K}[t]$ de uma matriz $0 \neq A \in \mathbb{M}_{n\times n}(\mathbb{K})$ é o polinômio mônico e de menor grau tal que $m(A) = 0$.

Lema 5.2.7 *Existe o polinômio mínimo de uma matriz $A \neq 0$.*

Demonstração: O espaço $\mathbb{M}_{n\times n}(\mathbb{K})$ é um espaço vetorial de dimensão n^2. Assim, as n^2+1 matrizes $I,\ A,\ A^2, \ldots, A^{n^2}$ são linearmente dependentes. Quer dizer, existem escalares não todos nulos $a_0, a_1, \ldots, a_{n^2} \in \mathbb{K}$, tais que

$$a_0 I + a_1 A + \ldots + a_{n^2} A^{n^2} = 0.$$

Definindo $p(t) = a_0 + a_1 t + \ldots + a_{n^2} t^{n^2}$, temos $0 \neq p \in \mathbb{K}[t]$ e $p(A) = 0$. Dividindo pelo coeficiente do termo de maior grau, podemos supor que p seja mônico. O polinômio mínimo então existe, como decorrência da aplicação do Princípio da Boa Ordenação ao conjunto de todos os polinômios mônicos que anulam A.

$\square$

Lema 5.2.8 *Se $p \in \mathbb{K}[t]$ e $p(A) = 0$, então p é múltiplo de m.*

Demonstração: Se $\mathcal{I}$ denota o conjunto de todos os polinômios $p \in \mathbb{K}[t]$ tais que $p(A) = 0$, claramente a soma de dois polinômios em $\mathcal{I}$, bem como qualquer múltiplo de $p \in \mathcal{I}$ estão em $\mathcal{I}$ (quer dizer, $\mathcal{I}$ é um ideal). A divisão euclidiana de p por m nos dá $p = qm + r$. Como $r = p - qm$ pertence a $\mathcal{I}$, concluímos que $r = 0$.

$\square$

5.3 O teorema de Cayley-Hamilton

Nessa seção apresentaremos um dos resultados mais importantes da Álgebra Linear[4]:

Teorema 5.3.1 (Cayley-Hamilton)
Seja V um espaço vetorial de dimensão n. Se $p \in \mathbb{K}[t]$ é o polinômio característico de $T : V \to V$, então $p(T) = 0$.

Demonstração: Seja $0 \neq v \in V$ arbitrário. Queremos mostrar que $p(T)v = 0$. Seja m o maior natural tal que o conjunto

$$S = \{v, Tv, \ldots, T^{m-1}v\}$$

é linearmente independente. Então

$$T^m v = \alpha_0 v + \ldots + \alpha_{m-1} T^{m-1} v. \tag{5.5}$$

[4] Veja a elegante demonstração apresentada no livro de Lax.

Seja $W = < S >$. Então os elementos de S formam uma base de W. Afirmamos que $T(W) \subset W$. De fato, se $w \in W$, então $w = \beta_0 v + \beta_1 Tv + \ldots + \beta_{m-1}T^{m-1}v$, para escalares $\beta_0, \ldots, \beta_{m-1} \in \mathbb{K}$. Assim, Assim,

$$Tw = \beta_0 Tv + \beta_1 T^2 v + \ldots + \beta_{m-1}T^m v.$$

A igualdade (5.5) garante o afirmado.

Seja T_i a restrição de T ao subespaço W. A representação de T_i na base S é

$$A = \begin{pmatrix} 0 & 0 & \cdots & 0 & \alpha_0 \\ 1 & 0 & \cdots & 0 & \alpha_1 \\ 0 & 1 & \cdots & 0 & \alpha_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \alpha_{m-1} \end{pmatrix}.$$

Logo,

$$\begin{aligned} \det(tI - A) &= \det \begin{pmatrix} t & 0 & \cdots & 0 & -\alpha_0 \\ -1 & t & \cdots & 0 & -\alpha_1 \\ 0 & -1 & \cdots & 0 & -\alpha_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & t-\alpha_{m-1} \end{pmatrix} \\ &= t \det \begin{pmatrix} t & 0 & \cdots & -\alpha_1 \\ -1 & t & \cdots & -\alpha_2 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & \cdots & -1 & t-\alpha_{m-1} \end{pmatrix} + \\ &\quad (-\alpha_0)(-1)^{m+1} \det \begin{pmatrix} -1 & t & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 \end{pmatrix}. \end{aligned}$$

Como o determinante da última matriz é $(-1)^{m-1}$, o último termo é justamente $-\alpha_0$. Procedendo do mesmo modo, obtemos

$$\det(tI - A) = t^m - \alpha_{m-1}t^{m-1} - \ldots - \alpha_0 = p_W(t),$$

sendo $p_W(t)$ o polinômio característico de T restrito a W. Mas a equação (5.5) mostra então que $p_W(T)v = 0$.

Afirmamos agora que $p(t) = q(t)p_W(t)$, para algum polinômio $q(t)$. Daí decorre o resultado, pois $v \neq 0$ foi escolhido arbitrariamente e $p(T)v = q(T)p_W(T)v = 0$. Para provar a afirmação, basta notar se completarmos S de forma a obter uma base $\mathcal{B}$ de V, a representação de T nessa base é

$$\begin{pmatrix} A & B \\ 0 & C \end{pmatrix}.$$

O resultado então decorre do exercício 8 do capítulo 4, pois

$$\det(tI - T) = \det(tI - A)\det(tI - C) = p_W(t)q(t)$$

(em cada expressão, as ordens das matrizes I são diferentes).

□

5.4 O teorema da decomposição primária

Vamos agora generalizar o que fizemos na primeira seção desse capítulo. Suponhamos a seguinte decomposição em soma direta:

$$V = W_1 \oplus W_2 \oplus \cdots \oplus W_j,$$

com $\dim W_i = n_i$ e $T(W_i) \subset W_i$, para $1 \leq i \leq j$. Seja $\mathcal{B}_i = \{w_{i1}, \ldots, w_{ik_i}\}$ uma base de W_i. Então

$$\mathcal{B} = \{w_{11}, \ldots, w_{1k_1}, w_{21}, \ldots, w_{2k_2}, \ldots, w_{j1}, \ldots, w_{jk_j}\}$$

é uma base de V, de acordo com o corolário 5.1.6. Uma vez que $T(W_i) \subset W_i$ temos que

$$T(w_{i\ell}) = c_{1\ell}w_{i1} + c_{2\ell}w_{i2} + \cdots + c_{k_i\ell}w_{ik_i}.$$

Assim, a representação de T na base $\mathcal{B}$ é

$$T_{\mathcal{B}} = \begin{pmatrix} T_1 & 0 & \cdots & 0 \\ 0 & T_2 & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & T_j \end{pmatrix},$$

em que T_i é um bloco de tamanho $k_i \times k_i$:

$$T_i = \begin{pmatrix} c_{11} & c_{21} & \cdots & c_{1k_i} \\ c_{21} & c_{22} & \cdots & c_{2k_i} \\ \vdots & \vdots & \ddots & \vdots \\ c_{k_i1} & c_{k_i2} & \cdots & c_{k_ik_i} \end{pmatrix}.$$

(Veja exemplo 5.4.5, abaixo).

Definição 5.4.1 *Seja $p \in \mathbb{K}[t]$ o polinômio característico da aplicação linear $T : V \to V$, em que V é um espaço vetorial de dimensão finita n. Suponhamos que*

$$p(t) = [p_1(t)]^{s_1} \cdots [p_j(t)]^{s_j}$$

seja a decomposição de p em fatores irredutíveis, com $p_i \neq p_k$ para $i \neq k$. Definimos, para $i = 1, \ldots, j$, o **autoespaço generalizado** *associado ao polinômio p_i como o conjunto de todos os vetores $v \in V$ para os quais existe um inteiro positivo k tal que*

$$[p_i(T)]^k v = 0.$$

No caso em que $p_i(t) = t - \lambda_i$, sendo λ_i um autovalor de T, os elementos não-nulos do autoespaço generalizado são chamados **autovetores generalizados** *de T associados ao autovalor λ_i.*

Para $k \in \mathbb{N}^*$, seja $N_k(p_i)$ o núcleo de $[p_i(T)]^k$. Claramente temos que

$$N_1(p_i) \subset N_2(p_i) \subset \cdots .$$

Como $N_k(p_i)$ é um subespaço do espaço de dimensão finita V para todo $k \in \mathbb{N}$, esses subespaços precisam ser todos iguais à partir de certo índice $k \in \mathbb{N}$. Seja $d_i = d(p_i)$ o menor inteiro positivo com tal propriedade, isto é,

$$N_{d_i}(p_i) = N_{d_i+1}(p_i) = \cdots , \quad \text{mas} \quad N_{d_i-1}(p_i) \neq N_{d_i}(p_i).$$

O inteiro positivo d_i é chamado **índice** de $p_i(T)$.

Lema 5.4.2 *Os subespaços $N_k(p_i)$ são invariantes pelo operador T, para todo $k \in \mathbb{N}^*$. Se $W_i = \ker[p_i(T)]^{d_i}$, então o polinômio mínimo de T restrito a W_i é $[p_i(T)]^{d_i}$.*

Demonstração: Seja $w \in N_k(p_i) = \ker[p_i(T)]^k$. Então $[p_i(T)]^k Tw = T[p_i(T)]^k w = 0$, mostrando que $Tw \in N_k(p_i)$.

A afirmação sobre o polinômio mínimo decorre da definição de d_i.

□

Teorema 5.4.3 (Decomposição Primária)

Seja $T : V \to V$ uma aplicação linear e $p \in \mathbb{K}[t]$ seu polinômio característico. Se

$$p(t) = [p_1(t)]^{s_1} \cdots [p_j(t)]^{s_j}$$

é a decomposição de $p(t)$ em fatores irredutíveis, com $p_i \neq p_k$ para $i \neq k$, então, se d_i é o índice de $p_i(T)$, o polinômio mínimo de T é

$$m(t) = [p_1(t)]^{d_1} \cdots [p_j(t)]^{d_j},$$

em que $0 < d_i \leq s_i$ para $i = 1, \ldots, j$. Em outras palavras, o polinômio mínimo possui todos os fatores irredutíveis do polinômio característico de T. Além disso,

$$V = W_1 \oplus \cdots \oplus W_j,$$

em que $W_i = \ker[p_i(T)]^{d_i}$, com $T(W_i) \subset W_i$.

Demonstração: Seja $m \in \mathbb{K}[t]$ o polinômio mínimo de T. De acordo com o Teorema de Cayley-Hamilton 5.3.1 e o lema 5.2.8, os únicos fatores irredutíveis presentes na decomposição de m são fatores irredutíveis de p. Incluindo fatores irredutíveis $[p_i(t)]^0$ do polinômio característico p que eventualmente estejam ausentes na decomposição de m, podemos escrever

$$m(t) = m_1(t) \cdots m_j(t),$$

com $m_i(t) = [p_i(t)]^{r_i}$ e $r_i \geq 0$ para $i = 1, \ldots, j$. (Vamos mostrar que $r_i = d_i > 0$ para todo $i = 1, \ldots, j$).

Como $m(T) = 0$, vemos que todo vetor $v \in V$ pertence ao núcleo de $m(T) = m_1(T) \cdots m_j(T)$. Como os polinômios $m_1(t) = [p_1(t)]^{r_1}, \ldots, m_j(t) = [p_j(t)]^{r_j}$ são primos entre si dois a dois, podemos aplicar o corolário 5.2.6 e concluir que

$$V = N_{m_1 \cdots m_j} = N_{m_1} \oplus \cdots \oplus N_{m_j}. \tag{5.6}$$

Consideremos agora $q_i(t) := [p_i(t)]^{d_i}$. Pela definição de d_i, se $0 \leq r_i \leq d_i$, então $N_{m_i} \subset N_{q_i} = W_i$ e $V = N_{m_1 \ldots m_j} \subset N_{q_1 \ldots q_k}$. Assim, pelo corolário 5.2.6,

$$V = N_{q_1} \oplus \cdots \oplus N_{q_j} = W_1 \oplus \cdots \oplus W_j. \tag{5.7}$$

Se $r_i > d_i$ ainda temos $N_{m_i} \subset N_{q_i}$, pois a definição de d_i garante que $N_{q_i} = N_{m_i}$. Em outras palavras, a decomposição (5.7) sempre é válida e, tendo em conta o lema 5.4.2, provamos a decomposição afirmada no enunciado do teorema.

Vamos agora provar que $r_i = d_i$. Denotando $T_i = T|_{W_i}$, temos que $q_i(T_i) = 0$, pela definição de W_i. Assim $(q_1 \ldots q_j)T = 0$ e, como $m(t)$ é o polinômio mínimo de T, $m(t)$ divide $q_1(t) \ldots q_j(t)$ e portanto $r_i \leq d_i$. Mas a definição de d_i garante a existência de $x \in W_i$ tal que $x \notin [p_i(T)]^{r_i}$ para $r_i < d_i$. Como $N_{m_i} \subset N_{q_i}$, isso contradiz a existência das decomposições(5.6) e (5.7). Logo $r_i = d_i$.

□

Observação 5.4.4 No caso especial em que $\mathbb{K} = \mathbb{C}$, o Teorema da Decomposição Primária é conhecido como **Teorema Espectral**.

Levando em conta o Teorema de Cayley-Hamilton, vemos que $d_i \leq r_i$. ◀

Exemplo 5.4.5 Considere a aplicação $T : \mathbb{R}^5 \to \mathbb{R}^5$ definida por

$$T(x_1, x_2, x_3, x_4, x_5) = (10x_1 - 7x_4 + x_5, -x_3, x_2, 13x_1 - 9x_4 + x_5, 4x_1 - 3x_4 + x_5).$$

Procuramos uma base $\mathcal{B}$ na qual realiza-se a decomposição primária de T.

A representação de T na base canônica do $\mathbb{R}^5$ é a matriz

$$A = \begin{pmatrix} 10 & 0 & 0 & -7 & 1 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 13 & 0 & 0 & -9 & 1 \\ 4 & 0 & 0 & -3 & 1 \end{pmatrix}.$$

O polinômio característico de A é o

$$\det(A - \lambda I) = \begin{pmatrix} 10-\lambda & 0 & 0 & -7 & 1 \\ 0 & -\lambda & -1 & 0 & 0 \\ 0 & 1 & -\lambda & 0 & 0 \\ 13 & 0 & 0 & -9-\lambda & 1 \\ 4 & 0 & 0 & -3 & 1-\lambda \end{pmatrix}.$$

Desenvolvendo esse determinante com relação à segunda coluna, obtemos:

$$\det(A-\lambda I) = -\lambda \det \begin{pmatrix} 10-\lambda & 0 & -7 & 1 \\ 0 & -\lambda & 0 & 0 \\ 13 & 0 & -9-\lambda & 1 \\ 4 & 0 & -3 & 1-\lambda \end{pmatrix} - \det \begin{pmatrix} 10-\lambda & 0 & -7 & 1 \\ 0 & -1 & 0 & 0 \\ 13 & 0 & -9-\lambda & 1 \\ 4 & 0 & -3 & 1-\lambda \end{pmatrix}.$$

Desenvolvendo esses dois determinantes, obtemos

$$\begin{aligned} \det(A-\lambda I) &= \lambda^2 \det \begin{pmatrix} 10-\lambda & -7 & 1 \\ 13 & -9-\lambda & 1 \\ 4 & -3 & 1-\lambda \end{pmatrix} + \det \begin{pmatrix} 10-\lambda & -7 & 1 \\ 13 & -9-\lambda & 1 \\ 4 & -3 & 1-\lambda \end{pmatrix} \\ &= (\lambda^2+1)[\lambda^3 - 2\lambda^2 + \lambda] \\ &= \lambda(\lambda^2+1)(\lambda-1)^2. \end{aligned}$$

Pelo Teorema da Decomposição Primária,

$$\mathbb{R}^5 = \ker A \oplus \ker(A^2+I) \oplus \ker(A-I)^2.$$

Encontramos $\ker A$ resolvendo o sistema $Ax = 0$. Assim,

$$\begin{pmatrix} 10 & 0 & 0 & -7 & 1 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 13 & 0 & 0 & -9 & 1 \\ 4 & 0 & 0 & -3 & 1 \end{pmatrix} \to \begin{pmatrix} 10 & 0 & 0 & -7 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 3 & 0 & 0 & -2 & 0 \\ 4 & 0 & 0 & -3 & 1 \end{pmatrix}$$

Logo, $x_2 = x_3 = 0$, $x_4 = 3x_1/2$, $x_5 = -4x_1 + 3x_4 = -4x_1 + 9x_1/2 = x_1/2$. Assim, a solução geral de $Ax = 0$ é $x = (2x_1, 0, 0, 3x_1, x_1)$ e o vetor $v_1 \in \mathcal{B} = \{v_1, v_2, v_3, v_4, v_4\}$ pode ser escolhido como $v_1 = (2, 0, 0, 3, 1)$. Calculando $A^2 + I$ e resolvendo o sistema $(A^2 + I)x = 0$, encontramos a solução geral

$$(0, x_2, x_3, 0, 0),$$

de modo que os vetores v_2 e v_3 podem ser escolhidos como

$$v_2 = (0, 1, 0, 0, 0) \quad \text{e} \quad v_3 = (0, 0, 1, 0, 0).$$

Da mesma forma o sistema $(A - I)^2 x = 0$, cuja solução geral é

$$(x_1, 0, 0, x_4, 3x_1 - 2x_4)$$

o que nos permite escolher os vetores

$$v_4 = (1, 0, 0, 0, 3) \quad \text{e} \quad v_5 = (0, 0, 0, 1, -2).$$

Consideremos então a base $\mathcal{B} = \{v_1, v_2, v_3, v_4, v_5\}$. Vamos representar a aplicação linear T nessa base. Temos:

$$\begin{aligned} T(2,0,0,3,1) &= 0 = 0v_1 \\ T(0,1,0,0,0) &= (0,0,1,0,0) = 1v_3 \\ T(0,0,1,0,0) &= (0,-1,0,0,0) = -v_2 \\ T(1,0,0,0,3) &= (13,0,0,16,7) = 13v_4 + 16v_5 \\ T(0,0,0,1,-2) &= (-9,0,0,-11,-5) = -9v_4 - 11v_5 \end{aligned} .$$

Assim, a representação de T na base $\mathcal{B}$ é

$$T_{\mathcal{B}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 13 & -9 \\ 0 & 0 & 0 & 16 & -11 \end{pmatrix}.$$

A submatriz (0), (matriz 1×1) situada na extremidade superior esquerda corresponde à restrição de T ao subespaço invariante $\ker A$. A matriz

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

é a restrição de T ao subespaço invariante $\ker(A^2+I)$. A matriz

$$\begin{pmatrix} 13 & -9 \\ 16 & -11 \end{pmatrix}$$

é a restrição de T ao subespaço invariante $\ker(A-I)^2$. ◀

Proposição 5.4.6 *Com a notação do teorema* 5.4.3*, o subespaço* $W_i = \ker[p_i(T)]^{d_i}$ *tem dimensão igual ao grau de* $[p_i(t)]^{s_i}$*, em que* s_i *é a multiplicidade de* p_i *como fator irredutível do polinômio característico* $p(t)$*.*

Demonstração: Como o polinômio característico de uma matriz $n\times n$ tem grau n, basta mostrar que o polinômio característico de T restrito a W_i é justamente $[p_i(t)]^{s_i}$.

Seja $\mathcal{B}_i$ uma base de W_i. Como $V = W_1 \oplus \cdots W_j$, a representação de T na base $\mathcal{B}$ formada pelos vetores de cada base $\mathcal{B}_i$ é

$$T_{\mathcal{B}} = A = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_j \end{pmatrix},$$

em que A_i é um bloco de tamanho $k_i \times k_i$, em que k_i é a dimensão de W_i. Assim

$$\det(tI - A) = \det(tI - A_1)\cdots\det(tI - A_j). \tag{5.8}$$

Observe que $\det(tI - A_i)$ é o polinômio característico de T_i, a restrição de T ao subespaço W_i. Como o polinômio mínimo de T_i é $[p_i(t)]^{d_i}$ (pelo lema 5.4.2), o Teorema da Decomposição Primária 5.4.3 garante que o polinômio característico de T_i é uma potência de $p_i(t)$. Da igualdade (5.8) segue que o polinômio característico de T_i é $[p_i(t)]^{s_i}$.

□

Lema 5.4.7 *Sejam* $S, T : V \to V$ *duas aplicações lineares no espaço de dimensão finita* V*. Suponhamos que* $ST = TS$*. Então existe uma base de* V *na qual tanto* S *como* T *realizam sua decomposição primária.*

Demonstração: Como no Teorema da Decomposição Primária, seja $W_i = \ker[p_i(T)]^{d_i}$. Se $w_i \in W_i$, afirmamos que $Sw_i \in W_i$ (isto é, que W_i é um subespaço invariante também para S). De fato,

$$[p_i(T)]^{d_i} S w_i = S[p_i(T)]^{d_i} w_i = 0.$$

Isso mostra o afirmado.

□

No caso de $\mathbb{K} = \mathbb{C}$ podemos obter um resultado mais forte:

Proposição 5.4.8 *Sejam $S, T : V \to V$ duas aplicações lineares no espaço de dimensão finita V sobre $\mathbb{C}$. Suponhamos que $ST = TS$. Então existe uma base de V formada por autovetores generalizados de S e T.*

Demonstração: Já vimos que $W_i = \ker[p_i(T)]^{d_i}$ é invariante por S. Todos os elementos não-nulos de W_i são, por definição, autovetores generalizados de T. Aplicamos então o Teorema da Decomposição Primária ao subespaço W_i com respeito a B e obteremos uma divisão desse subespaço em subespaços formados por autovetores generalizados de B.

□

Note que a demonstração anterior mostra que a proposição 5.4.8 permanece válida para qualquer número de operadores que comutam. Mais precisamente,

Proposição 5.4.9 *Se $T_1, \ldots, T_m : V \to V$ são aplicações lineares no espaço de dimensão finita V sobre $\mathbb{C}$ e se $T_iT_j = T_jT_i$ para $i, j \in 1, \ldots, m$, então existe uma base de V formada por autovetores generalizados para todas as aplicações $T_1, \ldots, T_m$.*

5.5 A forma canônica de Jordan

Seja V um espaço vetorial de dimensão finita. Nessa seção mostraremos como encontrar uma base de V na qual um operador linear $T : V \to V$ assume uma matriz especialmente simples.

Definição 5.5.1 *Uma matriz complexa J, $n \times n$, está na* **forma canônica de Jordan** *se*

$$J = \begin{pmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_k \end{pmatrix}, \quad \text{em que} \quad J_i = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & \cdots & 0 & \lambda \end{pmatrix},$$

em que λ é um dos autovalores distintos $\lambda_1, \ldots, \lambda_j$ da matriz J. (Ao autovalor λ_i pode estar associado mais do que um bloco J_i; às vezes se define J_i com a sub-diagonal de 1's situando-se abaixo da diagonal principal).

Mostraremos a seguir que toda matriz complexa é semelhante a uma matriz na forma canônica de Jordan. Lembramos que, no caso de matrizes reais, sempre podemos vê-la como uma matriz complexa, de modo que, nesse sentido, o resultado abaixo é geral. Mais do que isso, a necessidade de considerarmos o corpo complexo é para garantir que os autovalores estão todos presentes no corpo (veja exemplo 5.5.4, abaixo).

Teorema 5.5.2 (Jordan)
Sejam $A, B \in \mathbb{M}_{n\times n}(\mathbb{C})$ duas matrizes semelhantes, isto é,

$$A = P^{-1}BP.$$

Então

(*i*) *A e B possuem os mesmos autovalores λ_i;*

(*ii*) *os espaços $N_j(\lambda_i) = \ker(A-\lambda_i I)^j$ e $M_j(\lambda_i) = \ker(B-\lambda_i I)^j$ possuem a mesma dimensão para todo $j \in \mathbb{N}$ e todo autovalor λ_i.*

Reciprocamente, se estas duas condições se verificam, então A e B são semelhantes.

Demonstração: (" $\Rightarrow$ ") Notamos inicialmente que os núcleos de duas matrizes semelhantes têm dimensão igual. De fato, se $C = Q^{-1}DQ$ e $\{x_1, \ldots, x_k\}$ é uma base do núcleo de C, então $\{Qx_1, \ldots, Qx_k\}$ é uma base do núcleo de D.

Temos também que se A e B são semelhantes, então também são as matrizes $A-aI$ e $B-aI$, bem como qualquer potência delas:

$$(A-aI)^m = P^{-1}(B-aI)^m P.$$

A relação (*ii*) decorre então dos núcleos dessas matrizes terem a mesma dimensão. Em particular, como um elemento $v \in \ker(A-\lambda_i I)^{d_i} \setminus \ker(A-\lambda_i I)^{d_i-1}$ é tal que $(A-\lambda_i I)^{d_i-1}v$ é um autovetor de A associado ao autovalor λ_i, (*i*) decorre de (*ii*).

Para mostrarmos a recíproca, denotaremos $N_k = \ker(A-\lambda_i I)^k$. Começamos pelo

Lema 5.5.3 *A aplicação*

$$A-\lambda_i I : \frac{N_{k+1}}{N_k} \to W_i$$

tem imagem contida em $\frac{N_k}{N_{k-1}}$ e é injetiva.

Demonstração: Seja $x \in \frac{N_{k+1}}{N_k}$. Isso quer dizer que $(A-\lambda_i I)^{k+1}x = 0$ e $(A-\lambda_i I)^k x \neq 0$. Consideremos então $(A-\lambda_i I)x$. Como $(A-\lambda_i I)^k(A-\lambda_i I)x = (A-\lambda_i I)^{k+1}x$, vemos que $(A-\lambda_i I)x \in N_k$. Por outro lado, $(A-\lambda_i I)^{k-1}(A-\lambda_i I)x = (A-\lambda_i I)^k x \neq 0$, mostrando que $(A-\lambda_i I)x \notin N_{k-1}$.

Afirmamos agora que essa aplicação é injetiva. Sejam $x, y \in \frac{N_{k+1}}{N_k}$, com $(A-\lambda_i I)x = (A-\lambda_i I)y$. Então $(A-\lambda_i I)(x-y) = 0$, o que é um absurdo, pois então $x-y$ estaria em N_k.

□

Vamos agora construir uma base especial para W_i. Lembramos que uma base de N_k/N_{k-1} é obtida ao se escolher uma base para N_{k-1} e então completá-la para uma base de N_k; os elementos introduzidos formam a base procurada. (Veja o Teorema 1.4.4).

Seja $x_1, \ldots, x_\ell$ uma base de N_{d_i}/N_{d_i-1}. De acordo com o lema, os elementos

$$(A-\lambda_i I)x_1, \ldots, (A-\lambda_i I)x_\ell$$

são linearmente independentes e pertencem a N_{d_i-1}/N_{d_i-2}. Completamos esses elementos até obter uma base desse espaço. Pelo mesmo raciocínio, a imagem por $(A-\lambda_i I)$ dos elementos dessa base é linearmente independente e podemos novamente completar esse conjunto até obter uma base; procedemos desse modo até chegarmos ao espaço N_1. A base de W_i assim construída é a **base de Jordan** do subespaço W_i. Obtemos então uma base do espaço inteiro ao obtermos as bases de Jordan de cada espaço W_i. Essa base é chamada **base de Jordan**.

De acordo com a hipótese (ii), os subespaços $M_k = \ker(B-\lambda_i)^k$ têm a mesma dimensão do espaço correspondente N_k. Em outras palavras, o procedimento aplicado a N_k, se repetido para a matriz B, produzirá o mesmo número de elementos para cada base de M_k/M_{k-1}. Em outras palavras, existe uma aplicação P que faz corresponder a cada elemento da base de Jordan $x_j \in N_k/N_{k-1}$ o elemento correspondente na base de $y_j \in M_k/M_{k-1}$, de modo que $(A-\lambda_i I)x_j$ seja levado em $(B-\lambda_i I)y_j$. Conhecida a imagem dos vetores da base, existe uma única aplicação linear que estende essa aplicação; seja P tal extensão. Como base está sendo levada em base, essa aplicação linear tem inversa. O mesmo procedimento aplicado ao autoespaço associado a λ_i constrói a aplicação P.

Finalmente, a definição de P garante que $P(A-\lambda_i I)x_j = (B-\lambda_i I)y_j = (B-\lambda_i I)Px_j$. Assim, $P(A-\lambda_i I) = (B-\lambda_i I)P$. Decorre daí que $PA = BP$, como desejado.

$\square$

Exemplo 5.5.4 Seja $T : \mathbb{R}^4 \to \mathbb{R}^4$ definido por

$$T(x_1,x_2,x_3,x_4) = (2x_1 - x_2 + x_4, 3x_2 - x_3, x_2 + x_3, -x_2 + 3x_4).$$

Vamos obter a forma canônica de Jordan de T, bem como a base na qual T assume essa forma.

O polinômio característico de T de $p(t) = (t-3)(t-2)^3$ (verifique!). Assim, todos os autovalores de T estão no corpo $\mathbb{R}$ e podemos obter a forma de Jordan de T. Para o autovalor 3, a forma escalonada reduzida de

$$(T-3I) = \begin{pmatrix} -1 & -1 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \quad \text{é} \quad \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Assim, o subespaço $W_1 = \ker(T-3I)$ do Teorema da Decomposição Primária é dado por

$$\{(x_1,0,0,x_1);\ x_1 \in \mathbb{R}\}.$$

Da mesma forma se verifica que

$$\begin{aligned} \ker(T-2I) &= \{(x_1,x_2,x_2,x_2); x_1,x_2 \in \mathbb{R}\} \\ \ker(T-2I)^2 &= \{(x_1,x_2+x_3,2x_3,2x_2);\ x_1,x_2,x_3 \in \mathbb{R}\}. \end{aligned}$$

Como a dimensão de $\ker(T-2I)^3$ é igual à multiplicidade de 2 como raiz do polinômio característico $p(t)$ de T , temos que o espaço W_2 do Teorema da Decomposição Primária é dado por $\ker(T-2I)^2$ (veja proposição 5.4.6).

O subespaço W_1 tem base $(1,0,0,1) = w_1$. Esse é o primeiro elemento da base de Jordan (ele é responsável por um bloco 1×1 no Teorema da Decomposição Primária).

Agora vamos obter a base de Jordan de W_2. Para isso, começamos por obter um vetor em $W_2 = N_2 = \ker(T - 2I)^2$ que **não** está em $N_1 = \ker(T - 2I)$ (é possível obter apenas um vetor assim, pois a diferença de dimensão entre esses espaços é 1). Ele fornecerá a base de N_2/N_1 da demonstração do Teorema de Jordan. Ora, claramente o vetor $w_4 = (0, 1, 0, 2) \in N_2$ e $w_4 \notin N_1$. Calculamos então $w_3 = (T - 2I)w_4 = (1, 1, 1, 1)$. (A demonstração do Teorema de Jordan garante que $w_3 \in N_1$ e que w_3, w_4 são linearmente independentes). Para obtermos uma base de N_1, escolhemos o vetor $w_2 = (1, 0, 0, 0) \in N_1$, que claramente é linearmente independente com w_3. (Mais uma vez, a demonstração do Teorema de Jordan garante que $\{w_2, w_3, w_4\}$ são linearmente independentes).

Temos assim a base $\mathcal{B} = \{w_1, w_2, w_3, w_4\}$, que é a base de Jordan de T. Os vetores w_2 e w_3 são autovetores de T associados ao autovalor 2 (pois eles pertencem a N_1). Finalmente, $(T - 2I)w_4 = w_3$, de modo que $Tw_4 = 2w_4 + w_3$. Assim, representando T na base $\mathcal{B}$, obtemos

$$T_{\mathcal{B}} = J = \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix},$$

que é a forma canônica de Jordan de T. ◀

Observação 5.5.5 Comparando com o exemplo acima, se tivéssemos encontrado dois vetores distintos para N_2/N_1, consideraríamos o ciclo formado pelo primeiro e então o ciclo formado pelo segundo e ordenaríamos a base nessa ordem. ◀

Exemplo 5.5.6 Obtenha uma base $\mathcal{B}$ na qual a matriz A esteja na forma canônica de Jordan:

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

O polinômio característico de A é $p(t) = (t - 2)^5(t + 1)$, pois a matriz A é triangular superior (enuncie e demonstre esse resultado sobre matrizes triangulares).

Se chamamos de W_1 o subespaço relacionado ao autovalor -1, vemos que $\dim W_1 = 1$ e que uma base para esse subespaço é dado pelo vetor e_6. (Você consegue obter essa base sem fazer qualquer conta?). Denotaremos $v_1 = e_6$ o primeiro vetor da base procurada.

Consideremos agora o espaço W_2, associado ao autovalor 2. Temos que $\dim W_2 = 5$ e que

$$A - 2I = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -3 \end{pmatrix}.$$

Se chamamos de $N_1 = \ker(A - 2I)$, vemos que $\dim N_1 = 2$. (Você consegue perceber isso sem fazer qualquer conta? Lembre-se que o número de linhas nulas no escalonamento de $A - 2I$ fornece os graus de liberdade nas soluções de $(A - 2I)x = 0$, isto é, a dimensão desse espaço.)

Temos

$$\begin{aligned}
N_1 = \ker(A - 2I) &= \{(0, 0, x_3, -x_3, x_4, x_4/3);\ x_3, x_4 \in \mathbb{R}\}\\
N_2 = \ker(A - 2I)^2 &= \{(0, 0, x_3, x_4, x_5, (3x_5 - x_4 - x_3)/9)\}\\
N_3 = \ker(A - 2I)^3 &= \{(0, x_2, x_3, x_4, x_5, (-2x_2 - 3x_3 - 3x_4 + 9x_5)/27\}\\
N_4 = \ker(A - 2I)^4 &= \{(x_1, x_2, x_3, x_4, x_5, (27x_5 - 9x_4 - 9x_3 - 6x_2 - 10x_1)/81\}
\end{aligned}$$

Uma vez que $\dim \ker(A - 2I)^4 = 5$, vemos que o coeficiente que estabiliza o espaço $A - 2I$ é 4. Se W_2 é o autoespaço generalizado associado ao autovalor 2, temos $W_2 = N_4$.

Claramente o vetor $v_6 = (1, 0, 0, 0, 0, -10/81) \in N_4 \setminus N_3$. Uma vez que a aplicação

$$A - 2I : \frac{N_4}{N_3} \to \frac{N_3}{N_2}$$

é injetiva e $\dim N_3/N_2 = 1$ (pois existe em N_3 apenas um grau de liberdade a mais do que em N_2, temos que

$$v_5 = (A - 2I)v_6 = (0, 1, -1, 0, 1, 10/27)$$

é o quinto vetor da base procurada.

Pelo mesmo motivo, $v_4 = (A - 2I)^2 v_6 = (A - 2I)v_5 = (0, 0, 0, 1, 0, -1/9)$ é o quarto vetor da base procurada (note que $v_4 \in N_2/N_1$).

O terceiro vetor da base é $v_3 = (A - 2I)^3 v_6 = (A - 2I)v_4 = (0, 0, 0, 0, 1, 1/3)$ é um autovetor de A, pois ele pertence a N_1. Como N_1 tem dimensão 2, existe um outro vetor nesse espaço, linearmente independente com v_3. Esse é o vetor $v_2 = (0, 0, 1, -1, 0, 0)$.

Tendo obtido os vetores $\{v_1, \ldots, v_6\}$, a representação de A nessa base é dada por

$$\begin{aligned}
Av_1 &= -v_1 \quad (\text{pois } (A + I)v_1 = 0)\\
Av_2 &= 2v_2 \quad (\text{pois } (A - 2I)v_2 = 0)\\
Av_3 &= 2v_3 \quad (\text{pois } (A - 2I)v_3 = 0)\\
Av_4 &= v_3 + 2v_4 \quad (\text{pois } (A - 2I)v_4 = v_3)\\
Av_5 &= v_4 + 2v_5 \quad (\text{pois } (A - 2I)v_5 = v_4)\\
Av_6 &= v_5 + 2v_6 \quad (\text{pois } (A - 2I)v_6 = v_5)
\end{aligned}$$

A representação de A nessa base é

$$J = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0\\ 0 & 2 & 0 & 0 & 0 & 0\\ 0 & 0 & 2 & 1 & 0 & 0\\ 0 & 0 & 0 & 2 & 1 & 0\\ 0 & 0 & 0 & 0 & 2 & 1\\ 0 & 0 & 0 & 0 & 0 & 2 \end{pmatrix}.$$

A matriz J tem um bloco 1×1 associado ao autovalor -1. Associado ao autovalor 2 ela tem dois blocos de Jordan: o bloco 1×1 associado ao autovetor v_2 e o bloco 4×4 associado aos elementos $\{v_3, v_4, v_5, v_6\} = \{(A - 2I)^3 v_6, (A - 2I)^2 v_6, (A - 2I)v_6, v_6\}$. ◀

Exemplo 5.5.7 Seja

$$A = \begin{pmatrix} -1 & 1 & -1 & -3 & -1 & 7 \\ 0 & -1 & 1 & 2 & 3 & 2 \\ 0 & 0 & -1 & 0 & -2 & 1 \\ 0 & 0 & 0 & -1 & 1 & -2 \\ 0 & 0 & 0 & 0 & -1 & 3 \\ 0 & 0 & 0 & 0 & 0 & -4 \end{pmatrix},$$

cujo polinômio característico é (obviamente) $p(t) = (t+1)^5(t+4)$. Temos

$$\begin{aligned} \ker(A+4I) &= \{(-2x_1, 0, -x_1, x_1, -x_1, x_1) : x_1 \in \mathbb{R}\} \\ \ker(A+I) &= \{(x_1, x_2, -2x_2, x_2, 0, 0) : x_1, x_2 \in \mathbb{R}\} \\ \ker(A+I)^2 &= \{(x_1, x_2, -2x_3 - 2x_4, x_3, x_4, 0) : x_1, x_2, x_3, x_4 \in \mathbb{R}\} \\ \ker(A+I)^3 &= \{(x_1, x_2, x_3, x_4, x_5, 0) : x_1, x_2, x_3, x_4, x_5 \in \mathbb{R}\} \end{aligned}$$

Escolhemos $v_1 = (-2, 0, -1, 1, -1, 1) \in \ker(A+4I)$. Esse é o primeiro vetor de uma base na qual A é representada por sua forma canônica de Jordan.

Claramente, $v_6 = (0,0,1,0,0,0) \in \ker(A+I)^3 \setminus \ker(A+I)^2$. Seja então $(A+I)v_6 = v_5 = (-1,1,0,0,0,0)$ e $(A+I)v_5 = v_4 = (1,0,0,0,0,0) \in \ker(A+I)$. Como $\dim \frac{\ker(A+I)^2}{\ker(A+I)} = 2$, existe mais um vetor nesse espaço, linearmente independente com v_5. A primeira vista, poderíamos escolher o vetor $v = (0,1,0,0,0,0)$, pois ele está em $\ker(A+I)^2$ e não está em $\ker(A+I)$. Entretanto, em $\frac{\ker(A+I)^2}{\ker(A+I)}$, os vetores v_5 e v são linearmente dependentes: basta notar que a diferença entre eles é um vetor em $\ker(A+I)$. Uma escolha correta para o vetor de $\frac{\ker(A+I)^2}{\ker(A+I)}$, linearmente independente com v_5 é $v_3 = (0,0,-2,0,1,0)$ (verifique). Então $(A+I)v_3 = v_2 = (1,1,-2,1,0,0)$.

Notamos, em particular, que pode ser complicada a escolha de três vetores linearmente independentes num espaço quociente N_i/N_{i-1}. (Em geral, isso pode ser obtido por simples inspeção: o vetor v_4 escolhido acima tem uma coordenada que não está presente no espaço $\ker(A+I)$). Se essa inspeção não é suficiente, a melhor maneira é pensar como é construída a base do espaço N_i/N_{i-1}: partindo de uma base de N_{i-1} os elementos que completam a base de N_i formam a base do quociente. Esse é o processo computacional adequado quando a dimensão do quociente for grande. ◀

Teorema 5.5.8 *Toda matriz $A \in \mathbb{M}_{n\times n}(\mathbb{C})$ é semelhante à sua transposta.*

Demonstração: Uma vez que $\det A = \det A^T$, obtemos que o polinômio característico dessas duas matrizes é igual. Em particular, elas têm os mesmos autovalores.

Notamos que se q é um polinômio e B uma matriz $n \times n$, então $[q(B)]^T = q(B^T)$ (basta tomar a transposta). Se λ_i é um autovalor de A (e, portanto, de A^T), aplicando esse resultado para os polinômios $(t-\lambda_i)^k$ e então considerando a dimensão de seus núcleos, decorre do Corolário 3.4.5que a condição (ii) do Teorema de Jordan também é cumprida.

□

Teorema 5.5.9 *Um operador linear $T : V \to V$ é diagonalizável se, e somente se, o seu polinômio mínimo é produto de fatores lineares distintos.*

Demonstração: Suponhamos que T seja diagonalizável e $\lambda_1, \ldots, \lambda_k$ os autovalores distintos de T. Então V possui uma base formada por autovetores de T, de acordo com o teorema 5.1.4. Considere o polinômio

$$h(z) = (t - \lambda_1) \ldots (t - \lambda_k).$$

Se v é um autovetor de T associado ao autovalor λ_i, então $(T - \lambda_i I)v = 0$. Isso implica que $h(T)v = 0$ para qualquer autovetor de T. Como o Teorema da Decomposição Primária implica que o polinômio mínimo e característico possuem os mesmos fatores irredutíveis, mostramos que h é o polinômio mínimo de T.

Reciprocamente, se $p(t) = (t - \lambda_1) \ldots (t - \lambda_k)$ é o polinômio mínimo de T, então $W_i = \ker(T - \lambda_i I)$. Claramente todo elemento de W_i é um autovetor de T. Tomando bases $\mathcal{B}_i$ de cada espaço W_i, temos que $\mathcal{B} = \{\mathcal{B}_1, \ldots, \mathcal{B}_k\}$ é uma base de V formada por autovetores de T.

$\square$

5.6 A forma de Jordan real

Definição 5.6.1 *Sejam $A \in \mathbb{M}_{n\times n}$ e $z \in \mathbb{K}^n$ um vetor qualquer. Definimos $\overline{A} \in \mathbb{M}_{n\times n}$ como a matriz obtida ao se tomar o conjugado em cada uma das entradas de A e $\bar{z} \in \mathbb{K}^n$ como o vetor obtido ao se tomar o conjugado em cada uma das coordenadas de z.*

É de verificação imediata que $\overline{A + \lambda B} = \overline{A} + \bar{\lambda}\overline{B}$, $\overline{AB} = \overline{A}\,\overline{B}$ para quaisquer matrizes $A, B \in \mathbb{M}_{n\times n}$ e $\lambda \in \mathbb{K}$. Além disso, também vale $\overline{Az} = \overline{A}\bar{z}$ para qualquer $z \in \mathbb{K}^n$.

Definição 5.6.2 *Seja V um espaço vetorial real. Definimos a* **complexificação** *de V como sendo o conjunto*

$$V_{\mathbb{C}} = \{u + iv;\ u, v \in V\}.$$

Em $V_{\mathbb{C}}$ soma-se e multiplica-se por escalar (complexo) de maneira "natural". É fácil verificar que $V_{\mathbb{C}}$ torna-se, assim, um espaço vetorial sobre os complexos.

Seja $T : V \to V$ uma aplicação linear. Definimos a **complexificação** *de T como sendo a aplicação $T_{\mathbb{C}} : V_{\mathbb{C}} \to V_{\mathbb{C}}$ definida por $T_{\mathbb{C}}(u + iv) = Tu + iTv$.*

Se identificarmos o vetor $v \in V$ com o vetor $v + i0 \in V_{\mathbb{C}}$, V é um subespaço de $V_{\mathbb{C}}$. Essa identificação será usada no próximo resultado:

Lema 5.6.3 *Sejam V um espaço vetorial de dimensão finita e $T : V \to V$ uma aplicação linear. As seguintes afirmativas são válidas:*

(i) toda base de V é base de $V_{\mathbb{C}}$;

(ii) os polinômios característicos de T e $T_{\mathbb{C}}$ são iguais;

(iii) se λ é um autovalor de $T_{\mathbb{C}}$, então $\bar{\lambda}$ é também um autovalor de $T_{\mathbb{C}}$; as multiplicidades algébricas dos autovalores λ e $\bar{\lambda}$ são iguais;

(*iv*) *seja* $\tilde{W}$ *um subespaço tal que* $w = u + iv \in \tilde{W}$ *implica que* $\bar{w} = u - iv \in \tilde{W}$*. Então* $\tilde{W}$ *possui uma base formada por vetores reais.*

Demonstração: (i) Basta notar que as partes real u e imaginária v de qualquer vetor $u + iv$ podem ser escritas como combinação linear dos elementos da base de V.

(ii) Decorre imediatamente de (i) com a identificação $V \ni v = v + i0 \in V_{\mathbb{C}}$, pois então as representações de T e $T_{\mathbb{C}}$ numa base de V são iguais.

(iii) Sejam λ um autovalor de $T_{\mathbb{C}}$ e $p(z)$ o polinômio característico de $T_{\mathbb{C}}$. Como $p(z)$ também é o polinômio característico de T, os coeficientes de $p(z)$ são reais. Tomando o conjugado na equação $p(\lambda) = 0$, obtemos $p(\bar{\lambda}) = 0$, o que mostra que $\bar{\lambda}$ também é uma raiz do polinômio característico de $T_{\mathbb{C}}$. Se $p'(\lambda) = \ldots = p^{(d-1)}(\lambda) = 0$ e $p^{(d)}(\lambda) \neq 0$ (isto é, se λ é raiz de multiplicidade d do polinômio característico[5]), tomando o conjugado em cada uma dessas equações obtemos $p'(\bar{\lambda}) = \ldots = p^{(d-1)}(\bar{\lambda}) = 0$ e $p^{(d)}(\bar{\lambda}) \neq 0$, mostrando que $\bar{\lambda}$ também tem multiplicidade d.

(iv) Seja $\{w_1, \ldots, w_k\}$ uma base de $\tilde{W}$, com $w_j = u_j + iv_j$, $j = 1, \ldots, k$. Somando e subtraindo os vetores w_j e $\bar{w}_j$, obtemos que $u_j = u_j + i0$ e $v_j = v_j + i0$ estão em $\tilde{W}$. Assim, o conjunto $S = \{u_1, v_1, \ldots, u_k, v_k\}$ é um conjunto de vetores reais que gera $\tilde{W}$. Uma base formada de vetores reais é obtida ao se tomar um subconjunto de S com k elementos que seja linearmente independente em $V_{\mathbb{C}}$. □

Lema 5.6.4 *Sejam* $T : V \to V$ *um operador linear e* $T_{\mathbb{C}}$ *sua complexificação. Se o subespaço* $\tilde{W} \subset V_{\mathbb{C}}$ *possui uma base formada por vetores reais, então ele é a complexificação de um subespaço* $W \subset V$*.*

Demonstração: Todo vetor de $\tilde{W}$ é da forma $w = u + iv$, sendo u e v vetores reais. Escrevendo u e v em termos dos vetores da base real, segue imediatamente que $\tilde{W}$ é a complexificação do espaço real W gerado pelos vetores dessa base. □

Teorema 5.6.5 (Forma de Jordan real)

Seja $T : V \to V$ *um operador linear real. Então existe uma base* $\mathcal{C}$ *de* V *na qual* T *é representado por uma matriz* J*, diagonal em blocos, cujos blocos diagonais, além daqueles associados a autovalores reais e que são como na definição da forma de Jordan complexa, também podem ter a forma*

$$J_{\alpha,\beta} = \begin{pmatrix} D_{\alpha,\beta} & I_2 & 0 & \cdots & 0 \\ 0 & D_{\alpha,\beta} & I_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ldots & \vdots \\ 0 & 0 & 0 & D_{\alpha,\beta} & I_2 \\ 0 & 0 & 0 & 0 & D_{\alpha,\beta} \end{pmatrix} \quad \text{em que} \quad D_{\alpha,\beta} = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix},$$

sendo $\alpha + i\beta$ *um autovalor complexo de* $T_{\mathbb{C}}$ *e* I_2 *a matriz identidade* 2×2*.*

[5] Veja exercício 6.

Demonstração: De acordo com o Teorema da Decomposição Primária, o espaço vetorial V pode ser decomposto como soma direta de espaços invariantes pela aplicação T. Se $\lambda \in \mathbb{R}$ é um autovalor de T, obtemos o espaço invariante W_λ. A base do espaço W_λ na qual T assume sua forma de Jordan nesse espaço é então construída como na demonstração do teorema de Jordan 5.5.2.

Assim, podemos nos limitar ao caso de autovalores $\lambda \in \mathbb{C} \setminus \mathbb{R}$ da complexificação $T_\mathbb{C}$ de T. Suponhamos que $T_\mathbb{C}$ possua um autovalor $\lambda \notin \mathbb{R}$. Decorre do lema 5.6.3($iii$) que $\bar{\lambda}$ também é autovalor de $T_\mathbb{C}$, o que garante a existência dos espaços W_λ e $W_{\bar{\lambda}}$. Se os vetores $w_j = u_j + iv_j$ $(j = 1, \ldots, k)$ formam uma base de W_λ, temos que os vetores $u_j - iv_j$ formam uma base de $W_{\bar{\lambda}}$, de acordo com o exercício 13.

Afirmamos que

$$S = \{u_1, v_1, u_2, v_2, \ldots, u_k, v_k\}$$

é uma base de $W_\lambda \oplus W_{\bar{\lambda}}$ formada apenas por vetores reais. De fato, como $\dim W_\lambda = \dim W_{\bar{\lambda}} = k$, o conjunto S tem a dimensão do espaço $W_\lambda \oplus W_{\bar{\lambda}}$. Por outro lado, todo vetor desse espaço é combinação linear dos elementos de S. Isso mostra o afirmado.

Finalmente, se $w_1 = u_1 + iv_1$ satisfaz $T_\mathbb{C} w_1 = \lambda w_1$ para $\lambda = \alpha + i\beta \in \mathbb{C} \setminus \mathbb{R}$, então

$$T(u_1) + iT(v_1) = (\alpha u_1 - \beta v_1) + i(\beta u_1 + \alpha v_1).$$

Se, para $j \in \{2, \ldots, r\}$, temos $T_\mathbb{C} w_j = \lambda w_j + w_{j-1}$, vemos que

$$Tu_j + iTv_j = (\alpha u_j - \beta v_j + u_{j-1}) + i(\beta u_j + \alpha v_j + v_{j-1}),$$

de onde segue que, na base $\{u_1, v_1, u_2, v_2, \ldots, u_k, v_k\}$ de $W_\lambda \oplus W_{\bar{\lambda}}$, $T_\mathbb{C}$ é representado por bloco(s) da forma descrita no enunciado do teorema. Como $T_\mathbb{C} = T$ para qualquer dos vetores dessa base, a demonstração está completa.

$\square$

5.7 Exercícios

1. Ache o polinômio mínimo de $T : \mathbb{R}^3 \to \mathbb{R}^3$ dada por $T(x, y, z) = (3x + y - z, 2x + 2y - z, 2x + 2y)$.

2. Seja V um espaço vetorial de dimensão finita. Mostre que o polinômio característico e o polinômio mínimo da aplicação linear $T : V \to V$ independem da base $\mathcal{B}$ de V.

3. Sejam $T : \mathbb{K}^n \to \mathbb{K}^n$ uma aplicação linear e A a representação de T na base canônica do $\mathbb{K}^n$. Suponhamos que T possa ser representada por uma matriz diagonal na base $\mathcal{B} = \{v_1, \ldots, v_n\}$. Se P é matriz cujas colunas são as coordenadas de v_i (com relação à base canônica do $\mathbb{K}^n$), mostre que $D = P^{-1}AP$, em que D é a matriz diagonal cujos elementos diagonais são os autovalores de T.

4. Demonstre a Proposição 5.4.9.

5. Uma aplicação linear $N : V \to W$ é **nilpotente** se existe $m \in \mathbb{N}$ tal que $N^m v = 0$ para todo $v \in V$. Dê exemplo de uma matriz A, 2×2, tal que $A \neq 0$ e $A^2 = 0$. Ache um vetor $v \in \mathbb{R}^2$ tal que $\{v, Av\}$ seja uma base do $\mathbb{R}^2$ e que A coincida com sua representação nessa base. Considere agora uma matriz N, $n \times n$. Seja m_0 o menor de tais números $m \in \mathbb{N}$. Se $m_0 > 1$, mostre que existe $v \in V$ tal que

$$\{v, Nv, \ldots, N^{m_0-2}v, N^{m_0-1}v\}$$

é linearmente independente. Dê exemplo de uma matriz 4×4 tal que $A \neq 0$, $A^2 = 0$ e existam vetores $v, w \in \mathbb{R}^4$ tais que $\{v, Av, w, Aw\}$ seja uma base de $\mathbb{R}^4$ e A coincida com sua representação nessa base. Dê exemplo de uma matriz 5×5 tal que $A^2 \neq 0$ e $A^3 = 0$, para a qual existam vetores $v, w \in \mathbb{R}^5$ tais que $\{v, Av, w, Aw, A^2w\}$ seja uma base do $\mathbb{R}^5$ e A coincida com sua representação nessa base.

6. Suponha que o polinômio $p(z)$ seja da forma $(z - \lambda)^d q(z)$, com $q(\lambda) \neq 0$ e $d \in \{2, 3, \ldots\}$. Mostre que $p'(\lambda) = \ldots = p^{(d-1)}(\lambda) = 0$, mas $p^{(d)}(\lambda) \neq 0$. Dizemos então que a raiz λ de $p(z)$ tem **multiplicidade algébrica** d.

7. Encontre a decomposição primária da matriz

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}.$$

8. Leia o parágrafo 6.8 (Teorema da Decomposição Primária) do livro Álgebra Linear, de Hoffman e Kunze. Estude o exemplo 14 daquela seção.

9. Seja A uma matriz $n \times n$ diagonalizável. Se B comuta com A, mostre que A e B são simultaneamente diagonalizáveis.

10. Obtenha uma base $\mathcal{B}$ na qual as seguintes matrizes estejam na forma canônica de Jordan:

$$(i) \quad \begin{pmatrix} 2 & 5 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

$$(ii) \quad \begin{pmatrix} 1 & 1 & 0 & 0 & -1 & 0 & 4 & 0 \\ 0 & 1 & 1 & -1 & -1 & -3 & 3 & -4 \\ 0 & 0 & 1 & 0 & 1 & 1 & -2 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & -4 & -5 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & -5 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 \end{pmatrix}.$$

11. Sejam
$$m(t) = (t-\lambda_1)^{d_1} \dots (t-\lambda_r)^{d_r} \quad p(t) = (t-\lambda_1)^{s_1} \dots (t-\lambda_r)^{s_r}$$
os polinômios mínimo e característico do operador $T : V \to V$, sendo V um espaço sobre os complexos. Mostre que

    (i) existe ao menos um bloco $d_i \times d_i$ associado ao autovalor λ_i;

    (ii) o número de blocos associados ao autovalor λ_i é igual à multiplicidade geométrica de λ_i (isto é, à dimensão do espaço W_i associado ao autovalor λ_i). (O inteiro s_i é a multiplicidade algébrica do autovalor λ_i).

12. Determine todas as possíveis formas canônicas de Jordan para uma matriz

    (i) cujo polinômio característico é $p(t) = (t-2)^3(t-5)^2$;

    (ii) cujo polinômio mínimo é $m(t) = (t-2)^2$;

    (iii) cujo polinômio característico é $p(t) = (t-3)^4(t-5)^4$ e cujo polinômio mínimo é $m(t) = (t-3)^2(t-5)^2$.

    Sugestão: em cada caso, quais são as possíveis dimensões dos subespaços associados ao polinômio mínimo?

13. Definiremos a conjugada $\bar{B}$ de uma matriz B qualquer como a matriz obtida ao se tomar o conjugado em cada uma de suas entradas. Mostre que $\overline{B+C} = \overline{B} + \overline{C}$ e $\overline{BC} = \overline{BC}$. Em particular, se $u = (\alpha_1 \ \dots \ \alpha_n)^T$ é um vetor, $\bar{u} = (\overline{\alpha_1} \ \dots \ \overline{\alpha_n})^T$ é o **conjugado** do vetor u. Seja $T : V \to V$ um operador linear real e $T_{\mathbb{C}}$ sua complexificação. Mostre que, se $u + iv \in \ker(T_{\mathbb{C}} - \lambda I)^r$, então $u - iv \in \ker(T_{\mathbb{C}} - \bar{\lambda} I)^r$. Conclua que se $u_1 + iv_1, \dots, u_k + iv_k$ é uma base de W_λ, então $u_1 - iv_1, \dots, u_k - iv_k$ é uma base de $W_{\bar{\lambda}}$.

14. Verifique que a demonstração do teorema 5.6.5 garante, em particular, que os subespaços W_λ e $W_{\bar{\lambda}}$ associados aos autovalores conjugados $\lambda, \bar{\lambda}$ possuem a mesma dimensão. Você é capaz de dar uma outra demonstração desse fato?

15. Seja $T_{\mathbb{C}}$ a complexificação do operador $T : V \to V$, sendo V é um espaço vetorial real. Suponhamos que $\lambda \in \mathbb{R}$ seja um autovalor de $T_{\mathbb{C}}$ (e, portanto, de T). Mostre que, se $\{w_1, \dots, w_k\}$ é uma base do espaço invariante W_λ, com $w_j = u_j + iv_j$, então tanto $\{u_1, \dots, u_k\}$ quanto $\{v_1, \dots, v_k\}$ são bases de W_λ. Dê então uma demonstração do teorema 5.6.5 usando a complexificação $T_{\mathbb{C}}$ também para o caso de um autovalor $\lambda \in \mathbb{R}$.

    Suponhamos agora que $\lambda \in \mathbb{C} \setminus \mathbb{R}$ seja um autovalor de $T_{\mathbb{C}}$ e $\{w_1, \dots, w_k\}$ uma base de W_λ, sendo $w_j = u_j + iv_j$. É verdade que $\{u_1, \dots, u_k\}$ é uma base de W_λ?

16. Seja $T : \mathbb{R}^4 \to \mathbb{R}^4$ um operador que tem a forma de Jordan complexa dada por
$$\begin{pmatrix} i & 1 & 0 & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & -i & 0 \\ 0 & 0 & 0 & -i \end{pmatrix}.$$
Ache a sua forma de Jordan real.

Capítulo 6

Estrutura Euclidiana

6.1 Produto interno

Definição 6.1.1 *Seja V um espaço vetorial sobre o corpo $\mathbb{K}$. Um* **produto interno** *em V é uma aplicação $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{K}$ satisfazendo às seguintes propriedades:*

(i) $\langle v, w \rangle = \overline{\langle w, v \rangle}$;

(ii) $\langle u + \lambda v, w \rangle = \langle u, w \rangle + \lambda \langle v, w \rangle$;

(iii) $\langle v, v \rangle \geq 0$ *e* $\langle v, v \rangle = 0$ *se, e somente se,* $v = 0$.

Um espaço V com produto interno é chamado **euclidiano** *se ele tem dimensão finita*[1].

Exemplo 6.1.2 Se $V = \mathbb{R}^n$, o produto interno canônico (também chamado produto escalar) é definido por

$$\langle x, y \rangle = x \cdot y = \sum_{i=1}^{n} x_i y_i = (x_1 \ \ldots \ x_n) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},$$

em que $x = (x_1, \ldots, x_n)$ e $y = (y_1, \ldots, y_n)$.

Com a mesma notação, esse é um caso particular do produto interno canônico em $\mathbb{C}^n$, definido por

$$\langle x, y \rangle = x \cdot y = \sum_{i=1}^{n} x_i \overline{y_i} = (x_1 \ \ldots \ x_n) \begin{pmatrix} \overline{y_1} \\ \vdots \\ \overline{y_n} \end{pmatrix}.$$

◀

Observação 6.1.3 Se o espaço vetorial V possui uma base $\mathcal{B} = \{v_1, \ldots, v_n\}$, então

$$\langle x, y \rangle = [x]_{\mathcal{B}}^T [\overline{y}]_{\mathcal{B}}$$

define um produto interno em V.

[1]Em alguns textos, um espaço euclidiano é um espaço com produto interno, mesmo em dimensão infinita.

Assim, ao dizermos que um espaço vetorial é euclidiano, não estamos atribuindo uma propriedade especial a esse espaço. Estamos, na verdade, especificando que naquele espaço foi escolhido um determinado produto interno, entre os vários produtos internos com que ele poderia ser considerado.

Note que a definição do produto interno através de uma base é a generalização do exemplo 6.1.2. Veremos posteriormente uma certa recíproca a esse resultado, caracterizando produtos internos em espaços de dimensão finita. ◀

Definição 6.1.4 *Sejam u, v vetores do espaço com produto interno V. Esses vetores são* **ortogonais** (*ou* **perpendiculares**) *se $\langle u, v\rangle = 0$. Nesse caso escrevemos $u \perp v$.*

Posteriormente justificaremos geometricamente essa definição.

6.2 Norma

Definição 6.2.1 *Seja V um espaço vetorial sobre o corpo $\mathbb{K}$. Uma* **norma** *em V é uma aplicação $\|\cdot\| : V \to [0, \infty)$ satisfazendo às seguintes propriedades:*

(i) $\|v\| > 0$ *se* $v \neq 0$*;*

(ii) $\|\lambda v\| = |\lambda|\ \|v\|$, *para* $\lambda \in \mathbb{K}$*;*

(iii) $\|v + w\| \leq \|v\| + \|w\|$.

Se V possui uma norma, dizemos que V é um espaço **normado**.

O valor $\|v\|$ pode ser interpretado geometricamente como o comprimento do vetor v. Se $\|v\| = 1$, o vetor v é chamado **unitário**.

Seja V um espaço com produto interno. Consideremos (com abuso de notação) $\|v\| := \langle v, v\rangle^{1/2}$. Vamos mostrar que essa notação se justifica, isto é, que $\langle v, v\rangle^{1/2}$ realmente define uma norma. Começamos justificando a definição de perpendicularidade dada acima.

Teorema 6.2.2 (Pitágoras)
Seja V um espaço com produto interno e $\|x\| = \langle x, x\rangle^{1/2}$. Então, se $x \perp y$, temos

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2.$$

Demonstração: Basta desenvolver $\|x + y\|^2$:

$$\|x + y\|^2 = \langle x + y, x + y\rangle = \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle = \|x\|^2 + \|y\|^2,$$

pois x e y são ortogonais.

□

Suponhamos agora que V seja real. Então $\langle x + y, x + y\rangle = \|x\|^2 + 2\langle x, y\rangle + \|y\|^2$. Se vale o Teorema de Pitágoras, então $x \perp y$.

Proposição 6.2.3 (Desigualdade de Cauchy-Schwarz)
Seja V um espaço com produto interno. Então, se $\|x\| = \langle x, x\rangle^{1/2}$, temos para todos $x, y \in E$

$$|\langle x, y\rangle| \leq \|x\| \; \|y\|.$$

Demonstração: A prova que apresentaremos é bem geométrica (interprete!). Se $x = \lambda y$, então $|\langle x, y\rangle| = |\lambda| \; \langle y, y\rangle = |\lambda| \; \|y\|^2 = \|x\| \; \|y\|$. Se $x \neq \lambda y$, existe $\alpha \in \mathbb{K}$ tal que $|\langle y - \alpha x, x\rangle| = 0$. (De fato, basta tomar $\alpha := \langle y, x\rangle / \|x\|^2$; note que $\|x\| = 0$ está incluído no caso anterior). Então

$$0 \leq \|y - \alpha x\|^2 = \|y\|^2 - |\alpha|^2 \; \|x\|^2 = \|y\|^2 - \frac{|(y, x)|^2}{\|x\|^2},$$

donde obtemos a desigualdade de Cauchy-Schwarz.

□

Agora estamos em condições de justificar a notação $\|x\| = \langle x, x\rangle^{1/2}$.

Proposição 6.2.4 *Todo espaço com produto interno V tem uma norma definida por $\|x\| = \langle x, x\rangle^{1/2}$.*

Demonstração: A primeira propriedade de norma decorre imediatamente da definição do produto interno. Além disso,

$$\|\lambda v\|^2 = \langle \lambda v, \lambda v\rangle = \lambda \overline{\lambda} \langle v, v\rangle = |\lambda|^2 \, \|v\|^2.$$

Finalmente, temos que

$$\begin{aligned}
\|v + w\|^2 &= \langle v + w, v + w\rangle = \|v\|^2 + \langle v, w\rangle + \langle w, v\rangle + \|w\|^2 = \|v\|^2 + 2\mathcal{R}e \, \langle v, w\rangle + \|w\|^2 \\
&\leq \|v\|^2 + 2\mathcal{R}e \, |\langle v, w\rangle| + \|w\|^2 \\
&\leq \|v\|^2 + 2\|v\| \, \|w\| + \|w\|^2 = (\|v\| + \|w\|)^2
\end{aligned}$$

□

A seguinte propriedade de espaços com produto interno é imediata (desenvolva o lado esquerdo da equação):

Proposição 6.2.5 *Em todo espaço com produto interno vale a* **identidade do paralelogramo***:*

$$\|x + y\|^2 + \|x - y\|^2 = 2\left(\|x\|^2 + \|y\|^2\right).$$

6.3 Bases ortonormais

Definição 6.3.1 *Seja V um espaço com produto interno. Um conjunto $X \subset V$ é* **ortogonal** *se $u \perp v$ para quaisquer $u, v \in X$. Se, além disso, todos os seus vetores são unitários, então X é* **ortonormal**.

O próximo resultado vale até em espaços vetoriais de dimensão infinita:

Lema 6.3.2 *Todo conjunto ortogonal formado por vetores não nulos é linearmente independente.*

Demonstração: Sejam $x_1, \ldots, x_m \in X$ e suponhamos que

$$\alpha_1 x_1 + \ldots + \alpha_m x_m = 0.$$

Então,

$$0 = \langle 0, x_i \rangle = \langle \alpha_1 x_1 + \ldots + \alpha_m x_m, x_i \rangle = \alpha_1 \langle x_1, x_i \rangle + \ldots + \alpha_m \langle x_m, x_i \rangle = \alpha_i \langle x_i, x_i \rangle.$$

Como $\langle x_i, x_i \rangle = \|x_i\|^2 \neq 0$, temos $\alpha_i = 0$.

□

Proposição 6.3.3 *Seja $\{v_1, \ldots, v_n\}$ uma base ortonormal de um espaço euclidiano V. Então*

(i) $u = \langle u, v_1 \rangle v_1 + \ldots + \langle u, v_n \rangle v_n$;

(ii) $\|u\|^2 = |\langle u, v_1 \rangle|^2 + \ldots + |\langle u, v_n \rangle|^2$.

Demonstração: Se $u = \alpha_1 v_1 + \ldots + \alpha_n v_n$, então $\langle u, v_i \rangle = \alpha_i \langle v_i, v_i \rangle = \alpha_i$. Isso mostra (i).
Como

$$\|u\|^2 = \Big\langle \langle u, v_1 \rangle v_1 + \ldots + \langle u, v_n \rangle v_n, \langle u, v_1 \rangle v_1 + \ldots + \langle u, v_n \rangle v_n \Big\rangle = \sum_{i=1}^{n} \langle u, v_i \rangle \overline{\langle u, v_i \rangle} = \sum_{i=1}^{n} |\langle u, v_i \rangle|^2,$$

(ii) também se verifica.

□

Teorema 6.3.4 (Gram-Schmidt)
Dada uma base arbitrária $\{u_1, \ldots, u_n\}$ do espaço euclidiano V, existe uma base ortonormal $\{x_1, \ldots, x_n\}$ de V formada por vetores x_i que são combinações lineares dos vetores $u_1, \ldots, u_i$, para todo $i = 1, \ldots, n$.

Demonstração: Utilizaremos indução na dimensão do espaço, o caso $n = 1$ sendo trivial. Suponhamos construídos os vetores $x_1, \ldots, x_{k-1}$. Consideramos então

$$x_k = c \left(u_k - \sum_{i=1}^{k-1} c_i x_i \right).$$

Para obtermos x_k ortogonal a todos os x_i já escolhidos, basta definir $c_i = \langle u_k, x_i \rangle$ para $i = 1, \ldots, k-1$. Escolhemos então $1/c$ como a norma do vetor $u_k - \sum_{i=1}^{k-1} c_i x_i$. Note que $c > 0$.

□

Uma interpretação do teorema de Gram-Schmidt em termos de decomposição matricial será dada na seção 8.4. O teorema de Gram-Schmidt garante a existência de uma infinidade de bases ortonormais para espaços euclidianos. Dada uma base ortonormal $\{x_1, \ldots, x_n\}$, temos

$$x = a_1 x_1 + \ldots + a_n x_n.$$

Os escalares a_i podem ser facilmente determinados. Como a base é ortonormal, temos que

$$a_i = \langle x, x_i \rangle \quad i = 1, \ldots, n.$$

Consideremos então um outro vetor $y \in V$. Então

$$\langle x, y \rangle = \langle a_1 x_1 + \ldots + a_n x_n, b_1 x_1 + \ldots + b_n x_n \rangle = a_1 \overline{b_1} + \ldots + a_n \overline{b_n},$$

o que mostra que, com relação a uma base ortonormal[2], qualquer produto interno em V tem a forma dada pela observação 6.1.3. Em particular, quando $y = x$, temos

$$\|x\|^2 = a_1^2 + \ldots + a_n^2.$$

Podemos ainda explorar mais as relações acima com a observação 6.1.3. Se $x = a_1 x_1 + \ldots + a_n x_n$, concluímos facilmente que a aplicação

$$S : V \to \mathbb{K}^n, \quad Sx = (a_1, \ldots, a_n)$$

é um isomorfismo que transforma um dado produto interno em V no produto escalar usual no $\mathbb{K}^n$.

Notamos que, para $y \in V$ fixo, a aplicação $x \mapsto \langle x, y \rangle$ é uma aplicação linear. Reciprocamente, temos o importante

Teorema 6.3.5 (de representação de Riesz) *Toda aplicação linear $\ell : V \to \mathbb{K}$ num espaço euclidiano V pode ser escrita como um produto interno. Mais precisamente, existe um único $y \in V$ tal que*

$$\ell(x) = \langle x, y \rangle \quad \forall\, x \in V.$$

Compare o enunciado acima com o teorema 2.2.3. Existe uma generalização desse resultado para certos espaços com produto interno de dimensão infinita (os espaços de Hilbert).

Demonstração: Considere uma base ortonormal $x_1, \ldots, x_n \in V$. Se $x = \langle x, x_1 \rangle x_1 + \ldots + \langle x, x_n \rangle x$, então

$$\begin{aligned} \ell(x) &= \langle x, x_1 \rangle \ell(x_1) + \ldots + \langle x, x_n \rangle \ell(x_n) \\ &= \langle x, \overline{\ell(x_1)} x_1 \rangle + \ldots + \langle x, \overline{\ell(x_n)} x_n \rangle = \langle x, \overline{\ell(x_1)} x_1 + \ldots + \overline{\ell(x_n)} x_n \rangle. \end{aligned}$$

Defina $y = \overline{\ell(x_1)} x_1 + \ldots + \overline{\ell(x_n)} x_n$. A unicidade de y decorre de $\{x_1, \ldots, x_n\}$ ser uma base.

$\square$

Decorre então o seguinte

Corolário 6.3.6 *Se V é um espaço euclidiano real, a aplicação $\ell \mapsto y$ é um isomorfismo entre V' e V.*

[2] Veja o exercício 2.

6.4 Projeções ortogonais

Definição 6.4.1 *Seja $Y \subset V$ um subespaço do espaço com produto interno V. O* **complemento ortogonal** *de Y, denotado $Y^{\perp}$, é o conjunto*

$$Y^{\perp} = \{v \in V \ \langle v, y \rangle = 0 \ \forall \ y \in Y\}.$$

Claramente $Y^{\perp}$ é um subespaço de V.

Teorema 6.4.2 *Para qualquer subespaço $Y \subset V$ de um espaço euclidiano temos*

$$V = Y \oplus Y^{\perp}.$$

Além disso, vale

$$(Y^{\perp})^{\perp} = Y.$$

Demonstração: Seja $y \in Y \cap Y^{\perp}$. Então $\langle y, y \rangle = 0$ e, portanto, $y = 0$.

Seja $y_1, \ldots, y_m$ uma base ortonormal de Y e $v \in V$. Defina $z = v - \langle v, y_1 \rangle y_1 - \ldots - \langle v, y_m \rangle y_m$. Então $z \in Y^{\perp}$ e $v = y + z$, com $y = \langle v, y_1 \rangle + \ldots + \langle v, y_m \rangle y_m$.

Temos, por (i), $V = Y \oplus Y^{\perp}$ e também $V = Y^{\perp} \oplus (Y^{\perp})^{\perp}$. Daí decorre que (ii).

□

Observação 6.4.3 A demonstração acima continua válida para espaços de dimensão infinita, desde que $Y \subset V$ tenha dimensão finita. Uma outra demonstração é a seguinte: considere uma base ortogonal $\{y_1, \ldots, y_m\}$ de Y e então complete essa base de modo a obter uma base de V. O processo de Gram-Schmidt mostra que podemos completar com uma base ortogonal: $\{y_1, \ldots, y_m, w_1, \ldots, w_k\}$. Claramente temos que $Y^{\perp}$ é o espaço gerado por $\{w_1, \ldots, w_k\}$. ◀

Definição 6.4.4 *Na decomposição*

$$\begin{array}{rccc} V = & Y & \oplus & Y^{\perp} \\ v = & y & + & z, \end{array}$$

a componente y é chamada projeção ortogonal de v em Y e denotada $y = \pi_Y v$.

Teorema 6.4.5 *Sejam V um espaço euclidiano e $Y \subset V$ um subespaço. A aplicação $\pi_Y : V \to Y$ é linear e satisfaz $\pi_Y^2 = \pi_y$. A aplicação π_Y é a* **projeção ortogonal** *de V em Y.*

Demonstração: Seja $w \in W$ qualquer. O teorema 6.4.2 garante que $w = \bar{y} + \bar{z}$, com $\bar{y} \in Y$ e $\bar{z} \in Y^{\perp}$. Logo, $\lambda \bar{w} = \lambda \bar{y} + \lambda \bar{z}$. Assim, $x + \lambda w = (y + \lambda \bar{y}) + (z + \lambda \bar{z})$, o que mostra que $\pi_Y(x + \lambda w) = y + \lambda \bar{y} = \pi_Y x + \lambda \pi_Y w$.

Se $v = y + z$ é a decomposição de v, então $\pi_Y y = y$, mostrando que $\pi_Y^2 = \pi_Y$.

□

Teorema 6.4.6 *Seja Y um subespaço do espaço euclidiano V e $v \in V$. Entre todos os elementos $y \in Y$, aquele com menor distância até v é o elemento $\pi_Y v$.*

Demonstração: A decomposição dada pelo teorema 6.4.2 garante que $v - y = (\pi_Y v - y) + z$, com $z \in Y^{\perp}$. Pelo teorema de Pitágoras,

$$\|v - y\|^2 = \|\pi_Y v - y\|^2 + \|z\|^2.$$

Assim, $\|v - y\|$ é mínima quando $y = \pi_Y v$.

□

6.5 A adjunta de uma aplicação linear

Sejam V, W espaços euclidianos.

Proposição 6.5.1 *Dada uma aplicação linear $T : V \to W$, existe uma única aplicação linear $T^* : W \to V$, chamada* **adjunta** *de T, satisfazendo*

$$\langle Tv, w\rangle = \langle v, T^*w\rangle \quad \forall\ v \in V,\ w \in W.$$

Demonstração: Para $w \in W$ fixo, a aplicação $v \mapsto \langle Tv, w\rangle$ pertence ao dual de V. O teorema de representação de Riesz garante então que existe um único $y \in Y$ tal que

$$\langle Tv, w\rangle = \langle v, y\rangle.$$

Defina $T^*w = y$. Está assim definida, para cada $w \in W$, uma aplicação $T^* : W \to V$.

Sejam $u, w \in W$ e $\lambda \in \mathbb{K}$. Então,

$$\langle v, T^*(u + \lambda v)\rangle = \langle Tv, u + \lambda w\rangle = \langle Tv, u\rangle + \bar{\lambda}\langle Tv, w\rangle = \langle v, T^*u\rangle + \langle v, \lambda T^*w\rangle.$$

Daí decorre a linearidade de T^*. Se $S^* : W \to V$ fosse outra aplicação linear tal que $\langle Tv, w\rangle = \langle v, S^*w\rangle$, então $\langle v, T^*w - S^*w\rangle = 0$ para todo $v \in V$. Escolhendo $v = T^*w - S^*w$, concluímos que $T^* = S^*$.

□

Exemplo 6.5.2 Seja $T : \mathbb{R}^2 \to \mathbb{R}^2$ dada por $T(x, y) = (ax + by, cx + dy)$. A representação de T com relação à base canônica do $\mathbb{R}^2$ é a matriz

$$T_{\mathcal{E}} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Então

$$\begin{aligned} \langle T(x_1, y_1), (x_2, y_2)\rangle &= (ax_1 + by_1)x_2 + (cx_1 + dy_1)y_2 \\ &= (ax_2 + cy_2)x_1 + (bx_2 + dy_2)y_1 \\ &= \langle (x_1, y_1), (ax_2 + cy_2, bx_2 + cy_2)\rangle, \end{aligned}$$

de onde concluímos que

$$[T^*]_{\mathcal{E}} = \begin{pmatrix} a & c \\ b & d \end{pmatrix}.$$

Note que se $T : \mathbb{C}^2 \to \mathbb{C}^2$ é dada por $T(x, y) = (ax + by, cx + dy)$ para $a, b, c, d \in \mathbb{C}$, então a representação de sua adjunta com relação à base canônica seria a conjugada da transposta da representação de T com relação à base canônica (verifique!). ◀

Seja $\mathcal{B} = \{x_1, \ldots, x_n\}$ uma base ortonormal de V e $T : V \to V$ uma aplicação linear. É fácil verificar que, se $A = (a_{ij})$ é a representação de T na base $\mathcal{B}$, então $a_{ij} = \langle Tx_i, x_j\rangle$. (Veja o exercício 17).

Se $\mathcal{C}$ é uma base arbitrária de V e $T : V \to V$ uma aplicação linear, a relação entre $[T]_{\mathcal{C}}$ e $[T^*]_{\mathcal{C}}$ é mais complicado do que no exemplo acima.

Proposição 6.5.3 *Sejam V, W, U espaços euclidianos e $T, S : V \to W$ e $R : W \to U$ aplicações lineares e $\lambda \in \mathbb{K}$. Então vale:*

(*i*) $I^* = I$;

(*ii*) $(T + S)^* = T^* + S^*$;

(*iii*) $(\lambda T)^* = \bar{\lambda} T^*$;

(*iv*) $(RT)^* = T^* R^*$;

(*v*) $(T^*)^* = T$.

(*vi*) $(T^{-1})^* = (T^*)^{-1}$

Demonstração: As provas dos resultados afirmados são muito semelhantes. Faremos apenas algumas delas.

(*ii*) $\langle v, (S+T)^* w\rangle = \langle (S+T)v, w\rangle = \langle Sv, w\rangle + \langle Tv, w\rangle = \langle v, S^* w\rangle + \langle v, T^* w\rangle = \langle v, (S^* + T^*)w\rangle$. A unicidade da adjunta garante então que $(S + T)^* = S^* + T^*$.

(*v*) $\langle v, T^{**} v\rangle = \langle T^* w, v\rangle = \overline{\langle v, T^* v\rangle} = \overline{\langle Tv, w\rangle} = \langle w, Tv\rangle$. De novo, a unicidade da adjunta garante o afirmado.

□

Teorema 6.5.4 *Sejam V, W espaços euclidianos e $T : V \to W$ uma aplicação linear. Então vale:*

(*i*) $\ker T^* = (\mathcal{I}m\, T)^\perp$;

(*ii*) $\ker T = (\mathcal{I}m\, T^*)^\perp$;

(*iii*) $\mathcal{I}m\, T^* = (\ker T)^\perp$;

(*iv*) $\mathcal{I}m\, T = (\ker T^*)^\perp$;

(*v*) *posto de* $T =$ *posto de* T^*.

Demonstração: Também nesse caso as demonstrações são muito semelhantes. Temos:

(*i*) $w \in \ker T^* \Leftrightarrow T^* w = 0 \Leftrightarrow \langle v, T^* w\rangle = 0\ \forall\ v \in V \Leftrightarrow \langle Tv, w\rangle = 0\ \forall\ v \in V \Leftrightarrow w \perp \mathcal{I}m\, T$.

Do mesmo modo se mostra (*ii*).

(*iii*) Basta passar (*ii*) ao complementar ortogonal: $((\mathcal{I}m\, T^*)^\perp)^\perp = (\ker T)^\perp$. Similarmente para (*iv*).

Finalmente, temos $V = \dim \ker T +$ posto de $T = \dim(\mathcal{I}m\, T^*)^\perp +$ posto de T. Decorre daí que $\dim \mathcal{I}m\, T^* =$ posto de T, mostrando o afirmado.

□

Definição 6.5.5 *Uma aplicação $T : V \to V$ é chamada* **auto-adjunta** *se $T^* = T$.*

Note que, se $\mathcal{B}$ é uma base ortonormal de V, $[T^*]_{\mathcal{B}} = ([T]_{\mathcal{B}})^*$. Além disso, se $T_{\mathcal{B}} = A$ é auto-adjunta, $P^{-1}AP = B$ implica que B é auto-adjunta.

Uma matriz A é **auto-adjunta** se $A^* = A$. No caso real, isso equivale a $A^T = A$ e a matriz é **simétrica**.

6.6 Norma de uma aplicação linear

Seja $T : V \to V$ uma aplicação linear definida no espaço euclidiano V.

Definição 6.6.1 *Definimos*

$$\|T\| = \max_{x \neq 0} \frac{\|Tx\|}{\|x\|}.$$

Chamamos $\|T\|$ *de* **norma** *da aplicação linear* T.

Decorre imediatamente da definição que $\|Tx\| \leq \|T\| \, \|x\|$. O próximo resultado garante que $\| \cdot \|$ definida acima é uma norma no espaço vetorial $\mathcal{L}(V)$ de todas as aplicações lineares de V em V.

Proposição 6.6.2 *Seja* $T : V \to V$ *uma aplicação linear definida no espaço euclidiano* V. *Então*

(*i*) *A aplicação* $\| \cdot \| : \mathcal{L}(V) \to [0, \infty)$ *é uma norma;*

(*ii*) $\|ST\| \leq \|S\| \, \|T\|$.

Demonstração: Claramente $\|T\| \geq 0$ e $\|T\| = 0$ se, e somente se, $Tx = 0$ para todo $x \neq 0$. Além disso

$$\|\lambda T\| = \max_{x \neq 0} \frac{\|\lambda Tx\|}{\|x\|} = \max_{x \neq 0} \frac{|\lambda| \, \|Tx\|}{\|x\|} = |\lambda| \max_{x \neq 0} \frac{\|Tx\|}{\|x\|} = |\lambda| \, \|T\|.$$

Além disso,

$$\|S + T\| = \max_{x \neq 0} \frac{\|(S+T)x\|}{\|x\|} \leq \max_{x \neq 0} \frac{\|Sx\| + \|Tx\|}{\|x\|} \leq \max_{x \neq 0} \frac{\|Sx\|}{\|x\|} + \max_{x \neq 0} \frac{\|Tx\|}{\|x\|} = \|S\| + \|T\|.$$

(*ii*) $\|(ST)x\| = \|S(Tx)\| \leq \|S\| \, \|Tx\| \leq \|S\| \, \|T\| \, \|x\|$.

□

6.7 Isometrias

Definição 6.7.1 *Seja* $M : V \to V$ *uma aplicação (não necessariamente linear) definida no espaço euclidiano* V. *A aplicação* M *é uma* **isometria** *se, para quaisquer* $x, y \in V$, *temos*

$$\|Mx - My\| = \|x - y\|. \tag{6.1}$$

Decorre imediatamente da definição que a composta de duas isometrias é uma isometria. Um exemplo elementar de uma isometria é uma **translação**:

$$Tx = x + a$$

para $a \in V$ fixo. Dada uma isometria, podemos compô-la com uma translação e produzir assim uma isometria que leva $0 \in V$ em $0 \in V$. Reciprocamente, toda isometria é a composta de uma que leva $0 \in V$ no $0 \in V$ com uma translação.

Teorema 6.7.2 *Seja $M : V \to V$ uma isometria no espaço euclidiano V, com $M(0) = 0$. Então:*

(*i*) *Se V é um espaço sobre $\mathbb{R}$, então M é linear;*

Supondo adicionalmente que M seja linear no caso complexo, então vale:

(*ii*) *$M^*M = I$; reciprocamente, se essa igualdade é satisfeita, então M é uma isometria.*

(*iii*) *M possui inversa e sua inversa é uma isometria.*

(*iv*) *Se V é um espaço vetorial sobre $\mathbb{R}$, então $\det M = \pm 1$. No caso complexo, $|\det M| = 1$.*

Demonstração: (*i*) Tomando $y = 0$ em (6.1) vem que

$$\|Mx\| = \|x\|.$$

Vamos denotar $Mx = x'$, $My = y'$, etc. Temos então

$$\|x'\| = \|x\|, \quad \|y'\| = \|y\| \quad \text{e} \quad \|x' - y'\| = \|x - y\|, \tag{6.2}$$

de acordo com o que acabamos de mostrar e a equação (6.1). Elevando ao quadrado a última igualdade vem

$$\langle x', y' \rangle = \langle x, y \rangle, \tag{6.3}$$

mostrando que T preserva o produto interno. Desenvolvendo $\|z' - x' - y'\|^2$ em termos do produto interno, obtemos

$$\|z' - x' - y'\|^2 = \|z'\|^2 + \|y'\|^2 + \|x'\|^2 - 2\langle z', x' \rangle - 2\langle z', y' \rangle + 2\langle x', y' \rangle.$$

Do mesmo modo,

$$\|z - x - y\|^2 = \|z\|^2 + \|y\|^2 + \|x\|^2 - 2\langle z, x \rangle - 2\langle z, y \rangle + 2\langle x, y \rangle.$$

Decorre então de (6.2) e (6.3) que

$$\|z' - x' - y'\|^2 = \|z - x - y\|^2.$$

Escolhemos então $z = x + y$. O lado direito da igualdade acima é, então, nulo. Isso garante que o lado esquerdo também é, o que implica que $z' - x' - y' = 0$. Mas isso garante que $M(x + y) = Mx + My$. Mas também temos que

$$\langle M(\lambda x), My \rangle = \langle \lambda x, y \rangle = \lambda \langle x, y \rangle = \lambda \langle Mx, My \rangle = \langle \lambda Mx, My \rangle,$$

o que garante que $M(\lambda x) = \lambda Mx$. Isso mostra a linearidade de M no caso real.

Para mostrarmos (*ii*), partimos de (6.3) (note que essa expressão também vale para o caso complexo): a relação

$$\langle x, y \rangle = \langle Mx, My \rangle = \langle x, M^*My \rangle$$

é válida para quaisquer $x, y \in V$, de modo que

$$\langle x, M^*My - y \rangle = 0.$$

Escolhendo x como o termo no lado direito do produto interno, vemos que $M^*My = y$ para todo y. A recíproca é obtida invertendo a ordem na prova dada acima.

Decorre imediatamente de $M^*M = I$ que M tem inversa. Quando tomamos $x = M^{-1}y$ na igualdade $\|Mx\| = \|x\|$ vem $\|M^{-1}y\| = \|y\|$. Como M^{-1} é linear, essa aplicação é uma isometria. Note que $M^{-1} = M^*$.

Finalmente, no caso real, como $M^* = M^T$ e $\det M^T = \det M$, a igualdade $M^*M = I$ garante que $(\det M)^2 = 1$ e, portanto, $\det M = \pm 1$. No caso complexo, $M^* = \overline{M^T}$. Decorre daí que $\det M^* = \overline{\det M^T} = \overline{\det M}$. Assim, $\det M \overline{\det M} = 1$, provando o afirmado.

□

O significado geométrico de (iv) é que uma aplicação que preserva distâncias também preserva volumes. Uma aplicação linear M que satisfaz $M^*M = I$ é chamada **ortogonal** no caso real e **unitária** no caso complexo.

Como antes, uma matriz A é **ortogonal** (respectivamente, **unitária**) se $A^TA = I$ (resp., $A^*A = I$).

6.8 Exercícios

1. Seja V um espaço euclidiano complexo. Dê um exemplo mostrando que a validade do Teorema de Pitágoras não implica que $x \perp y$.

2. Seja $\{v_1, \ldots, v_n\}$ uma base arbitrária do espaço euclidiano real V. Defina $g_{ij} = \langle v_i, v_j \rangle$. Se $u = \alpha_1 v_1 + \ldots + \alpha_n v_n$ e $v = \beta_1 v_1 + \ldots + \beta_n v_n$, mostre que
$$\langle u, v \rangle = \sum_{i,j=1}^{n} g_{ij}\alpha_i\beta_j. \tag{6.4}$$
Mostre também que a matriz $G = (g_{ij})$ é simétrica e **positiva**, isto é, se $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$
$$\sum_{i,j=1}^{n} g_{ij}x_ix_j = x^TGx > 0$$
para todo $x \neq 0$.

   Reciprocamente, fixada a base $\{v_1, \ldots, v_n\}$ do espaço real V e dada uma matriz simétrica positiva G, mostre que (6.4) define um produto interno em V.

3. Seja V um espaço euclidiano complexo e $\{v_1, \ldots, v_n\}$m uma base desse espaço. Defina $g_{ij} = \langle v_i, v_j \rangle$ e mostre a equação 6.4. Verifique então que a matriz G é hermitiana (isto é, $\bar{G}^T = G$), e
$$x^TG\bar{x} > 0 \quad \forall\, 0 \neq x \in \mathbb{C}^n. \tag{6.5}$$
Reciprocamente, se G é uma matriz hermitiana e a equação (6.5) se verifica, então $\langle u, v \rangle = u^TG\bar{v}$ define um produto interno.

4. Seja $C([a,b],\mathbb{K})$ o espaço das funções contínuas $f : [a,b] \to \mathbb{K}$. Mostre que

$$\langle f, g\rangle := \int_a^b f(t)\overline{g(t)}dt$$

define um produto interno nesse espaço.

5. Considere agora o espaço $C([-\pi,\pi],\mathbb{R})$. Mostre que o conjunto

$$X := \{1, \operatorname{sen} t, \cos t, \operatorname{sen} 2t, \cos 2t, \ldots\}$$

é um conjunto ortogonal.

6. Considere então o espaço vetorial $C([-1,1],\mathbb{R})$. Seja $P \subset C([-1,1],\mathbb{R})$ o subespaço formado por todas as funções pares e $\mathcal{I} \subset C([-1,1],\mathbb{R})$ o subespaço formado por todas as funções ímpares. Mostre que $\mathcal{I} = P^{\perp}$.

7. Seja $\mathbb{R}[t]$ o espaço vetorial de todos os polinômios com coeficientes em $\mathbb{R}$. Verifique que

$$X = \{1, t, t^2, \ldots\}$$

é uma base desse espaço. Encontre os 4 primeiros termos da base obtida ao se aplicar o processo de ortogonalização de Gram-Schmidt à base X.

8. Seja V um espaço com produto interno. Mostre as **identidades de polarização**:

   (i) Se V é um espaço real, então

$$\langle u, v\rangle = \frac{1}{4}\|u+v\|^2 - \frac{1}{4}\|u-v\|^2.$$

   (ii) Se V é um espaço complexo, então

$$\langle u, v\rangle = \frac{1}{4}\|u+v\|^2 - \frac{1}{4}\|u-v\|^2 + \frac{i}{4}\|u+iv\|^2 - \frac{i}{4}\|u-iv\|^2.$$

9. Prove o corolário 6.3.6. O que acontece se V for um espaço complexo?

10. Seja $A = (a_{ij})$ uma matriz $m \times n$ real. Considere o sistema $Ax = b$, em que $b \in \mathbb{R}^m$. O sistema

$$A^T y = 0$$

é chamado **sistema homogêneo transposto**. Mostre que o sistema $Ax = b$ possui solução se, e somente se, b for ortogonal a qualquer solução do sistema homogêneo transposto. (Isso implica que podemos garantir a existência de soluções para o sistema $Ax = b$ sem necessitar exibir uma de suas soluções; basta verificar se b é ortogonal às soluções de $A^T y = 0$).

11. Seja $T : V \to V$ uma aplicação linear. Mostre que se $T^*T = 0$ então $T = 0$.

12. Seja V um espaço euclidiano complexo. Mostre que $T : V \to V$ é auto-adjunto se, e somente se, $\langle Tv, v\rangle \in \mathbb{R}$ para todo $v \in V$.

13. Seja V um espaço com produto interno e $\alpha, \beta \in V$ vetores fixos. Mostre que $Tv = \langle v, \alpha\rangle\beta$ define uma aplicação linear em V. Mostre que T^* existe e obtenha sua expressão.

14. Sejam W_1, W_2V subespaços do espaço com produto interno V. Mostre que $(W_1 + W_2)^\perp = W_1^\perp \cap W_2^\perp$ e $(W_1 \cap W_2)^\perp = W_1^\perp + W_2^\perp$

15. Seja $w_1, \ldots, w_m$ uma base ortonormal do subespaço W do espaço com produto interno V. Mostre que, para todo $v \in V$, vale a **desigualdade de Bessel**

$$\|v\|^2 \geq \sum_{j=1}^{m} |\langle v, w_j\rangle|^2.$$

16. Sejam $\mathcal{B} = \{v_1, \ldots, v_n\}$ e $\mathcal{C} = \{w_1, \ldots, w_m\}$ bases ortonormais dos espaços euclidianos V e W, respectivamente. Seja $T : V \to W$ uma aplicação linear. Mostre que

$$T_{\mathcal{B}}^{\mathcal{C}} = A = (a_{ij}), \quad \text{em que } a_{ij} = \langle T(v_j), w_i\rangle, \text{ para } i = 1, \ldots, m \text{ e } j = 1, \ldots, n.$$

Conclua que $(T^*)_{\mathcal{C}}^{\mathcal{B}} = B = (b_{ij})$, em que $b_{ij} = \overline{a_{ji}}$, generalizando assim o exemplo 7.6.

17. Sejam V, W espaços euclidianos. Dadas as aplicações $S, T \in \mathcal{L}(V, W)$, defina

$$\langle S, T\rangle = \operatorname{tr}(S^*T).$$

Mostre que está definido assim um produto interno em $\mathcal{L}(V, W)$. Se $A = (a_{ij}$ e $B = (b_{ij}$ são, respectivamente, as matrizes de S e T com relação a bases ortonormais de V e W, mostre que

$$\langle A, B\rangle = \sum_{i,j} a_{ij} b_{ij}.$$

18. Um **isomorfismo** dos espaços com produto interno V e W é uma bijeção linear $T : V \to W$ que satisfaz, adicionalmente, $\langle Tu, Tv\rangle = \langle u, v\rangle$, para todos $u, v \in V$ (isto é, T preserva o produto interno). Se $\dim V = \dim W$, mostre que as seguintes afirmações são equivalentes:

    (i) T preserva o produto interno;

    (ii) T é um isomorfismo (de espaços com produto interno);

    (iii) T leva toda base ortonormal de V em base ortonormal de W;

    (iv) T leva alguma base ortonormal de V em uma base ortonormal de W.

19. Sejam V e W espaços com produto interno. Mostre que $T : V \to W$ preserva produto interno se, e somente se, $\|Tv\| = \|v\|$ para todo $v \in V$ (T preserva norma).

20. Mostre que M é uma matriz unitária (respectivamente, ortogonal) se, e somente se, suas colunas formam uma base ortonormal de $\mathbb{K}^n$.

21. Seja $T : V \to V$ um operador auto-adjunto no espaço com produto interno V. Mostre que

    (i) $\|v + iTv\| = \|v - iTv\|$ para todo $v \in V$;

(ii) $v + iTv = u + iTu$ se, e somente se, $v = u$;

(iii) $\ker(I + iT) = \{0\}$;

(iv) $\ker(I - iT) = \{0\}$.

Mostre que se $\dim V = n$, então

$$U := (I - iT)(i + iT)^{-1}$$

é um operador unitário, chamado **transformada de Cayley** de T.

Capítulo 7

Teoria Espectral Euclidiana

7.1 Formas bilineares e quadráticas

Definição 7.1.1 *Seja V um espaço vetorial. Uma* **forma bilinear**[1] *é uma função $B : V \times V \to \mathbb{K}$ tal que, para quaisquer $\alpha \in \mathbb{K}$ e $u_1, u_2, v_1, v_2 \in V$,*

(i) $B(\alpha u_1 + u_2, v) = \alpha B(u_1, v) + B(u_2, v)$;

(ii) $B(u, \alpha v_1 + v_2) = \alpha B(u, v_1) + B(u, v_2)$.

Uma forma bilinear é **simétrica** *se $B(u, v) = B(v, u)$.*

Exemplo 7.1.2 Se V é um espaço euclidiano real, então $\langle \cdot, \cdot \rangle$ é uma forma bilinear simétrica.

Seja $A \in \mathbb{M}_{n\times n}(\mathbb{K})$ uma matriz. Definindo $B : \mathbb{K}^n \times \mathbb{K}^n \to \mathbb{K}$ por

$$B(u, v) = u^T A v = (u_1 \ u_2 \ \dots \ u_n) \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix},$$

obtemos uma forma bilinear. ◀

Denotaremos $\mathcal{L}_2(V)$ o conjunto das formas bilineares definidas em V. $\mathcal{L}_2(V)$ é um espaço vetorial com as definições usuais de soma de funções e multiplicação por escalar.

Vamos mostrar que todas as formas bilineares definidas em espaços euclidianos são como no exemplo 7.1.2.

Proposição 7.1.3 *Seja $\mathcal{B} = \{x_1, \dots, x_n\}$ uma base do espaço vetorial V sobre o corpo $\mathbb{K}$. Existe um isomorfismo entre $\mathcal{L}_2(V)$ e $\mathbb{M}_{n\times n}(\mathbb{K})$ tal que $B(u, v) = [u]_{\mathcal{B}}^T A [v]_{\mathcal{B}}$, em que $[u]_{\mathcal{B}}$ é a representação de u na base $\mathcal{B}$. A matriz A é chamada matriz de B na base $\mathcal{B}$.*

[1] O termo bilinear é um caso particular da denominação utilizada na definição 4.2.1.

Demonstração: Seja $\mathcal{B} = \{x_1, \ldots, x_n\}$ uma base de V, $u = u_1x_1 + \ldots + u_nx_n$ e $v = v_1x_1 + \ldots + v_nx_n$. Então

$$\begin{aligned} B(u,v) &= B(u_1x_1 + \ldots + u_nx_n, v_1x_1 + \ldots + v_nx_n) \\ &= u_1v_1B(x_1,x_1) + u_1v_2B(x_1,x_2) + \ldots + u_1v_nB(x_1,x_n) + \ldots \\ &\quad +u_nv_1B(x_n,x_1) + \ldots + u_nv_nB(x_n,x_n) \\ &= \sum_{i,j=1}^{n} u_iv_jB(x_i,x_j). \end{aligned} \tag{7.1}$$

Isso mostra que B fica completamente determinada pelos n^2 valores $B(x_i, x_j)$. Definimos então a matriz $A = (a_{ij})$ por $a_{ij} = B(x_i, x_j)$. Então

$$B(u,v) = (u_1 \ \ldots \ u_n) \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = [u]_{\mathcal{B}}^T A[v]_{\mathcal{B}}. \tag{7.2}$$

Assim, estabelecemos uma bijeção Ψ entre $\mathcal{L}_2(V)$ e $\mathbb{M}_{n\times n}(\mathbb{K})$, associando a matriz A à forma B. Temos que $\Psi(B_1 + \lambda B_2)$ é uma matriz $C = (c_{ij})$ tal que $c_{ij} = (B_1 + \lambda B_2)(x_i, x_j)$. Mas então $c_{ij} = B_1(x_1, x_j) + \lambda B_2(x_i, x_j) = A_1 + \lambda A_2$, em que $A_1 = \Psi(B_1)$ e $A_2 = \Psi(B_2)$, o que mostra que Ψ é linear.

□

Observação 7.1.4 Sejam $\mathcal{B}$ uma base do espaço euclidiano **real** V, e $\langle\cdot,\cdot\rangle$ o produto interno canônico do $\mathbb{R}^n$. Denotando o vetor $[w]_{\mathcal{B}} \in \mathbb{R}^n$ simplesmente por w, temos

$$u^TAv = \langle u, Av\rangle,$$

como se verifica facilmente (veja também o exercício 2 do capítulo 6). ◀

Observação 7.1.5 Pode-se mostrar que a derivada segunda de uma aplicação $f : \mathbb{R}^n \to \mathbb{R}$ é dada por uma forma bilinear $f''(x)$ (que varia com o ponto $x \in \mathbb{R}^n$). A proposição 7.1.3 garante que, para vetores $h, k \in \mathbb{R}^n$,

$$f''(x)(h,k) = h^TH_xk,$$

em que H_x é a matriz (que varia com o ponto x) **hessiana** de f. A forma quadrática q aparece no desenvolvimento de Taylor de f:

$$f(x+h) = f(x) + \frac{1}{1!}f'(x).h + \frac{1}{2!}q_x(h) + r(h),$$

em que

$$f'(x).h = \langle \nabla f(x), h\rangle$$

é o produto interno do gradiente de f em x por h,

$$q_x(h) = f''(x).(h,h) = h^TH_xh$$

e $r(h)$ denota o resto de Taylor. ◀

Definição 7.1.6 *Seja $B \in \mathcal{L}_2(V)$ uma forma bilinear sobre o espaço real V. A aplicação $q : V \to \mathbb{R}$ dada por $q(v) = B(v,v)$ chama-se* **forma quadrática**.

Assim, se $v = v_1x_1 + \ldots + v_nx_n$ é a expressão de v na base $\{x_1, \ldots, x_n\}$, de acordo com (7.1), toda forma quadrática $q(v)$ se escreve como

$$q(v) = \sum_{i=1}^{n} a_{ij}v_iv_j = \langle v, Av\rangle. \tag{7.3}$$

Quando trabalhamos com formas quadráticas, **podemos supor** que a matriz A seja simétrica: é claro que $(A + A^*)^* = A + A^*$; o operador

$$\frac{A + A^*}{2}$$

é chamada **parte auto-adjunta** do matriz A. Temos que

$$q(y) = \langle y, Ay\rangle = \left\langle y, \frac{A + A^*}{2}y\right\rangle.$$

7.2 Diagonalização de formas quadráticas

Dada uma forma quadrática (7.3) podemos, com uma mudança de coordenadas, obter que ela seja representada por uma matriz diagonal.

Para mostrar isso, introduziremos novas coordenadas:

$$Lv = z$$

em que L é uma matriz mudança de coordenadas.

Teorema 7.2.1 (Sylvester) *Dada uma forma quadrática* (7.3)*, é possível fazer uma mudança de coordenadas $Lv = z$ de modo que, na nova variável z, a forma quadrática q é diagonal, isto é,*

$$q(L^{-1}z) = \sum_{i=1}^{n} d_iz_i^2. \tag{7.4}$$

Existem muitas mudanças de variáveis que diagonalizam q. Entretanto, o número de termos positivos, negativos e nulos entre os coeficientes d_i é sempre o mesmo (**Lei da Inércia**)*.*

Demonstração: Seja $q(v) = \langle v, Av\rangle$, a matriz $A = (a_{ij})$ sendo simétrica. Se todos os termos a_{ij} são nulos, q já é diagonal. Suponhamos que todos os termos diagonais de q sejam nulos, mas que exista um termo a_{ij} diferente de zero, digamos $a_{12} = a_{21} \neq 0$. Os termos de q envolvendo v_1 e v_2 são

$$\begin{aligned} a_{12}v_1v_2 + a_{21}v_2v_1 + \left(\textstyle\sum_{j=3}^{n} a_{1j}v_1v_j\right) + \left(\textstyle\sum_{j=3}^{n} a_{j1}v_jv_1\right) + \left(\textstyle\sum_{j=3}^{n} a_{2j}v_2v_j\right) + \left(\textstyle\sum_{j=3}^{n} a_{j2}v_jv_2\right) \\ = 2a_{12}v_1v_2 + \left(\textstyle\sum_{j=3}^{n} a_{1j}2v_1v_j\right) + \left(\textstyle\sum_{j=3}^{n} a_{j2}2v_2v_j\right) \end{aligned}$$

Definimos então $w_1 = v_1 + v_2$ e $w_2 = v_1 - v_2$ e $w_k = v_k$ para $k = 3, \ldots, n$. (Assim, $2v_1 = w_1 + w_2$ e $2v_2 = w_1 - w_2$). Obtemos

$$\frac{a_{12}}{2}(w_1^2 - w_2^2) + \ldots,$$

mostrando assim que podemos supor, sem perda de generalidade, que q possua um termo diagonal diferente de zero.

Suponhamos então que $a_{11} \neq 0$. Agrupamos então os termos contendo v_1:

$$a_{11}v_1^2 + 2\sum_{j=2}^{n} a_{1j}v_1v_j.$$

Logo, podemos escrever esses termos como

$$a_{11}\left(v_1 + \frac{1}{a_{11}}\sum_{j=2}^{n} a_{1j}v_j\right)^2 - \frac{1}{a_{11}}\left(\sum_{j=2}^{n} a_{1j}v_j\right)^2$$

Definimos então a mudança de variável linear

$$z_1 = v_1 + \frac{1}{a_{11}}\sum_{j=2}^{n} a_{1j}v_j.$$

Podemos então escrever

$$q(v) = a_{11}z_1^2 + q_2(v),$$

em que q_2 é uma forma quadrática que depende apenas de $v_2, \ldots, v_n$. Tendo diagonalizado o termo em z_1, repetimos então o processo com a forma quadrática q_2.

Vamos agora mostrar a Lei da Inércia. Para isso, denotamos p_+, p_- e p_0 o número de termos positivos, negativos e nulos em (7.4).

Dizemos que q é **positiva** num subespaço $Y \subset V$ se $q(v) > 0$ para todo $0 \neq v \in Y$.

Afirmação: A dimensão do maior subespaço de V no qual q é positiva é p_+:

$$p_+ = \max \dim Y,\ q \text{ positiva em } Y.$$

Similarmente,

$$p_- = \max \dim Y,\ q \text{ negativa em } Y.$$

Para mostrarmos a afirmação, reordenamos os termos d_i de (7.4) de modo que os p primeiros sejam todos positivos, com $p = p_+$.

$$q(z) = d_1z_1^2 + \ldots + d_pz_p^2 + d_{p+1}z_{p+1}^2 + \ldots + d_nz_n^2. \tag{7.5}$$

Para $z = (z_1, \ldots, z_n) \in V$, seja S^+ o subespaço dos vetores da forma $(z_1, \ldots, z_p, 0, \ldots, 0)$. Claramente q é positiva em S^+. Isso mostra que $p^+ \leq \max \dim Y$, com q positiva em Y. Suponhamos que exista algum subespaço Y com q positiva em Y e $\dim Y > p$. Claramente $S^+ \subset Y$. Considere a aplicação $\pi : Y \to S^+$, $\pi(z) = \pi(z_1, \ldots, z_n) = (z_1, \ldots, z_p, 0, \ldots, 0)$. Como $\dim Y > \dim S^*$, existe $z \neq 0$ tal que $\pi(z) = 0$. Mas isso implica que as primeiras p componentes de z são nulas. Mas então, de acordo com (7.5), $q(z) \leq 0$, contradição. Analogamente se mostra a outra igualdade.

A afirmação garante então que os números p_+, p_- e p_0 podem ser definidos em termos de q, independente das coordenadas que colocam q na forma diagonal. Uma vez que $p_+ + p_- + p_0 = n$, isso completa a demonstração.

□

Teorema 7.2.2 *Dada uma matriz real simétrica, existe uma matriz invertível real M tal que*

$$M^*AM = M^TAM = D, \tag{7.6}$$

sendo D uma matriz diagonal.

Demonstração: Considerada a mudança de variáveis $Lv = z$ que diagonaliza a forma quadrática $q(v) = \langle v, Av\rangle$, seja $M = L^{-1}$. Então $v = Mz$ e

$$q(v) = \langle v, Av\rangle = \langle Mz, AMz\rangle = \langle z, M^*AMz\rangle.$$

Claramente q tem a forma (7.4) se, e somente se, M^*AM é uma matriz diagonal. Mostramos assim que os teoremas 7.2.1 e 7.2.2 são equivalentes.

□

7.3 Aplicações auto-adjuntas

Em muitas aplicações é importante utilizar mudanças de coordenadas tais que os comprimentos euclidianos da velha variável e da nova sejam o mesmo, isto é

$$\|v\|^2 = \|z\|^2.$$

Em termos da expressão matricial $v = Mz$, isso significa que M é uma isometria. Assim, de acordo com o teorema 6.7.2, M deve satisfazer $M^*M = I$.

Um dos resultados mais importantes da Matemática garante que, dada uma forma quadrática q, é possível diagonalizá-la por meio de uma mudança isométrica de coordenadas. Em outras palavras, que tanto (7.6) como $M^*M = I$ sejam satisfeitas.

Teorema 7.3.1 *Seja V um espaço euclidiano complexo e $H : V \to V$ uma aplicação auto-adjunta. Então os autovetores de H estão associados a autovalores reais e formam uma base ortonormal de V.*

Demonstração: De acordo com o Teorema da Decomposição Primária (especializado para o caso $\mathbb{K} = \mathbb{C}$), os autovetores generalizados de H geram o espaço V. Para mostrarmos o afirmado, precisamos mostrar que uma aplicação auto-adjunta satisfaz às seguintes propriedades adicionais:

(a) H possui apenas autovalores reais;

(b) H possui uma base formada por (autênticos) autovetores;

(c) Autovetores correspondentes a autovalores distintos são ortogonais.

De fato, uma vez mostrado $(b)-(c)$, podemos aplicar o processo de ortogonalização de Gram-Schmidt e obter bases ortonormais para os subespaços invariantes associados a cada autovalor. A afirmação (c) então implica que teremos uma base ortonormal formada por autovalores de T. (Assim, como conseqüência do teorema 5.1.4, H será representada nessa base por uma matriz diagonal).

(a) Seja v um autovetor associado ao autovalor λ de H. Então

$$\lambda\langle v,v\rangle = \langle\lambda v,v\rangle = \langle Hv,v\rangle = \langle v,Hv\rangle = \langle v,\lambda v\rangle = \overline{\lambda}\langle v,v\rangle,$$

de modo que $(\lambda-\overline{\lambda})\langle v,v\rangle = 0$. Isso mostra que $\lambda\in\mathbb{R}$, pois $\lambda=\overline{\lambda}$.

(b) Suponhamos que v seja um autovetor generalizado de H associado ao autovalor λ. Então $(H-\lambda I)^d v = 0$ para algum $d\in\mathbb{N}$. Queremos mostrar que $(H-\lambda I)v = 0$. Suponhamos inicialmente que $d=2$. Então, tomando o produto interno com v, obtemos

$$0 = \langle (H-\lambda I)^2 v, v\rangle = \langle (H-\lambda I)v, (H-\lambda I)v\rangle = \|(H-\lambda I)v\|^2.$$

Mas isso implica que $(H-\lambda I)v = 0$, como desejado.

Suponhamos agora que $d>2$. Reescrevemos $(H-\lambda I)^d v = 0$ como $(H-\lambda I)^2(H-\lambda I)^{d-2}v = 0$. Definindo $w = (H-\lambda I)^{d-2}v$, podemos concluir que $(H-\lambda I)w = 0$, ou seja, $(H-\lambda I)^{d-1}v = 0$. O resultado pode então ser mostrado por indução.

(c) Sejam $v,\ w$ autovetores associados aos autovalores distintos $\lambda,\ \mu\in\mathbb{R}$. Então

$$\lambda\langle v,w\rangle = \langle Hv,w\rangle = \langle v,Hw\rangle = \langle v,\mu w\rangle = \mu\langle v,w\rangle,$$

de modo que

$$(\lambda-\mu)\langle v,w\rangle = 0.$$

Como $\lambda\neq\mu$, isso implica $v\perp w$.

□

O próximo resultado não passa de uma reformulação do teorema 7.3.1 em termos de matrizes.

Teorema 7.3.2 *Seja H uma matriz complexa auto-adjunta. Então existem uma matriz unitária U e uma matriz diagonal D tais que*

$$U^*HU = D.$$

Demonstração: Decorre imediatamente dos exercícios 3 do capítulo 5 e 20 do capítulo 6

□

A versão para matrizes reais é a seguinte:

Teorema 7.3.3 *Seja H uma matriz real auto-adjunta (isto é, simétrica). Então existem uma matriz ortogonal P e uma matriz diagonal D tais que*

$$P^T HP = D.$$

Demonstração: Considerando a matriz H agindo sobre $\mathbb{C}^n$, a expressão

$$Hv = \lambda v$$

e o fato dos autovalores de H serem reais implicam que a parte real (e também a imaginária) de V também são autovetores de H. Assim, existe uma base ortonormal formada por autovetores reais de H. A demonstração então é como no teorema 7.3.2.

□

Observação 7.3.4 É possível dar uma demonstração alternativa do teorema 7.3.3, sem fazer uso dos exercícios utilizados naquela prova. Como já vimos, obtemos uma base ortonormal formada por autovetores (reais) de H. Seja $\mathcal{B} = \{v_1, \ldots, v_n\}$ essa base ortonormal. Se $v \in \mathbb{R}^n$, consideremos sua representação $z = (z_1\ z_2\ \ldots\ z_n)$ na base $\mathcal{B}$ (quer dizer, $z = [v]_{\mathcal{B}}$):

$$v = z_1 v_1 + z_2 v_2 + \ldots + z_n v_n. \tag{7.7}$$

Então

$$\|v\|^2 = \langle v, v\rangle = \sum_{i=1}^{n} z_i^2 = \|z\|^2. \tag{7.8}$$

Aplicando H em (7.7), obtemos

$$Hv = \lambda_1 z_1 v_1 + \ldots + \lambda_n z_n v_n, \tag{7.9}$$

em que λ_i é o autovalor associado ao autovetor v_i.

Substituindo (7.7) e (7.9) em $q(v) = \langle v, Hv\rangle$, vemos que

$$q(v) = \lambda_1 z_1^2 + \ldots + \lambda_n z_n^2.$$

Essa expressão mostra que a nova variável z diagonaliza a forma quadrática q. Combinando com o teorema 7.2.2, vemos que

$$P^* H P = D.$$

A equação (7.8) mostra que P é uma isometria e, portanto, $P^* = P^T = P^{-1}$. ◀

Vamos agora reescrever o teorema 7.3.1 em termos de projeções.

Teorema 7.3.5 (Resolução Espectral dos operadores auto-adjuntos)

Sejam V um espaço euclidiano e $H : V \to V$ uma aplicação linear auto-adjunta, com autovalores distintos $\lambda_1, \ldots, \lambda_k$. Seja E_{λ_j} o autoespaço associado ao autovalor λ_j, para $1 \le j \le k$. Se $\pi_j : V \to E_{\lambda_j}$ denota a projeção ortogonal sobre E_{λ_j}, então

$$I = \sum_{j=1}^{k} \pi_j.$$

e

$$H = \sum_{j=1}^{k} \lambda_j \pi_j.$$

As projeções ortogonais π_j satisfazem

$$\pi_i \pi_j = 0, \quad \textit{se } i \neq j, \quad \pi_j^2 = \pi_j \quad \textit{e} \quad \pi_j^* = \pi_j.$$

Demonstração: De acordo com o teorema 7.3.1 (ou teorema 7.3.3), vale

$$V = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_k},$$

em que os espaços E_{λ_j} são ortogonais dois a dois. Em outras palavras,

$$v = v_1 + \ldots + v_k, \quad v_j \in E_{\lambda_j}. \tag{7.10}$$

Definimos então $\pi_j(v) = v_j$. Claramente π_j é uma aplicação linear, satisfazendo $\pi_j^2 = \pi_j$ e $\pi_i \pi_j = 0$ se $i \neq j$. A expressão (7.10) pode ser escrita como

$$I = \sum_{j=1}^{k} \pi_j.$$

Aplicando o operador H em (7.10), como os elementos de E_{λ_j} são autovetores associados ao autovalor λ_j, obtemos

$$Hv = Hv_1 + \ldots + Hv_k = \lambda_1 v_1 + \ldots + \lambda_k v_k = \sum_{j=1}^{k} \lambda_j \pi_j(v).$$

Falta apenas mostrar que as projeções π_j são auto-adjuntas. Se $w = w_1 + \ldots + w_k$ com $w_j \in E_{\lambda_j}$, então

$$\langle \pi_j v, w \rangle = \left\langle v_j, \sum_{i=1}^{k} w_i \right\rangle = \sum_{i=1}^{k} \langle v_j, w_i \rangle = \langle v_j, w_j \rangle,$$

devido à ortogonalidade entre os espaços E_{λ_j}. Analogamente,

$$\langle v, \pi_j w \rangle = \langle v_j, w_j \rangle.$$

Isto mostra que

$$\langle \pi_j v, w \rangle = \langle v, \pi_j w \rangle.$$

□

O teorema 7.3.5 é especialmente útil para se definir funções de aplicações lineares auto-adjuntas. Por exemplo, decorre imediatamente daquele teorema que

$$H^2 = \sum_{j=1}^{k} \lambda_j^2 \pi_j$$

e, por indução,

$$H^m = \sum_{j=1}^{k} \lambda_j^m \pi_j.$$

Assim, para qualquer polinômio $q \in \mathbb{K}[t]$, temos

$$p(H) = \sum_{j=1}^{k} q(\lambda_j) \pi_j.$$

Definimos, para uma função f definida em $\mathbb{R}$,

$$f(H) = \sum_{j=1}^{k} f(\lambda_j)\pi_j.$$

Essa definição é conhecida como o **cálculo funcional** da aplicação auto-adjunta H. Por exemplo,

$$e^{Ht} = \sum_{j=1}^{k} e^{\lambda_j}\pi_j.$$

Uma outra conseqüência importante do teorema 7.3.5 diz respeito a aplicações auto-adjuntas que comutam:

Teorema 7.3.6 *Sejam $H, K : V \to V$ aplicações auto-adjuntas tais que*

$$HK = KH.$$

Então H e K podem ser simultaneamente diagonalizadas, isto é, existe uma base ortogonal de V formada por vetores que são, ao mesmo tempo, autovetores de K e H.

Demonstração: (Essa demonstração deve ser comparada com aquela da proposição 26 do capítulo de teoria espectral). Notamos inicialmente que $H - \lambda_j I$ comuta com K. Uma vez que

$$K(H - \lambda_j I)x = (H - \lambda_j I)Kx,$$

vemos que se $x \in \ker(H - \lambda_j I)$, então $Kx \in \ker(H - \lambda_j I)$.

Consideremos então a aplicação auto-adjunta $K : \ker(H - \lambda_j I) \to \ker(H - \lambda_j I)$ e aplicamos o teorema 7.3.1 (ou teorema 7.3.3). Obtemos então uma base de $\ker(H - \lambda_j I)$ formada por autovetores de K. Como todo elemento de $\ker(H - \lambda_j I)$ é um autovetor de H, obtivemos assim uma base ortogonal desse espaço formada por autovetores tanto de K quanto de H. Aplicamos então esse processo a cada autoespaço $\ker(H - \lambda_j I)$.

□

Note que o resultado anterior pode ser generalizado para qualquer número de aplicações auto-adjuntas que comutam duas a duas.

7.4 Aplicações normais

Definição 7.4.1 *Seja V um espaço euclidiano. Uma aplicação linear $T : V \to V$ é anti-auto-adjunta se*

$$T^* = -T.$$

No caso real, diz-se também que T é anti-simétrica.

De acordo com a proposição 6.5.3, temos

$$(iT)^* = -iT^* = iA,$$

mostrando que iT é uma aplicação auto-adjunta. Decorre imediatamente do teorema 7.3.1:

Teorema 7.4.2 *Seja $T : V \to V$ uma aplicação anti-auto-adjunta no espaço euclidiano complexo V. Então*

(i) *Os autovalores de T são imaginários puros;*

(ii) *Existe uma base ortonormal de V consistindo de autovetores de T.*

Demonstração: Considere uma base ortogonal $\{v_1, \ldots, v_j\}$ para iA. Então $(iT)v_j = \lambda_j v_j$, com $\lambda_j \in \mathbb{R}$. Logo

$$Tv_j = (-i\lambda)v_j,$$

mostrando que T tem os mesmos autovetores de iT, associados aos autovalores imaginários puros $(-i\lambda_j)$.

□

Introduzimos agora uma classe de aplicações lineares que engloba as aplicações auto-adjuntas, anti-auto-adjuntas e unitárias (ou ortogonais) como casos particulares.

Definição 7.4.3 *Seja V um espaço euclidiano. A aplicação linear $N : V \to V$ é* **normal** *se ela comuta com sua adjunta:*

$$NN^* = N^*N.$$

Teorema 7.4.4 *Uma aplicação linear normal $N : V \to V$ no espaço euclidiano complexo V possui uma base ortonormal consistindo de autovetores.*

Demonstração: Uma vez que N e N^* comutam, o mesmo acontece com

$$H := \frac{N + N^*}{2} \quad \text{e} \quad A := \frac{N - N^*}{2}.$$

As aplicações H e N são auto-adjunta e anti-auto-adjunta, respectivamente. Aplicamos então o teorema 7.3.5 às aplicações H e iA: existe uma base ortogonal formada por autovetores tanto de H quanto de iA e, assim, por autovetores tanto de H quanto de A. Como

$$N = H + A,$$

vemos que essa base é formada por autovetores de N. Note que, segundo os teoremas 7.3.1 e 7.4.2, se $Hv = av$ (com $a \in \mathbb{R}$) e $Av = (bi)v$ (com $b \in \mathbb{R}$), então $Nv = Hv + Av = (a + bi)v$.

□

O exercício 13 pede que se mostre a recíproca do resultado acima. Assim, existe uma base ortonormal $\mathcal{B}$ na qual $T_{\mathcal{B}}$ é diagonal se, e somente se, T é normal.

Teorema 7.4.5 *Uma aplicação linear normal $N : V \to V$ no espaço euclidiano real V possui uma base ortonormal $\mathcal{B}$ na qual N é uma matriz diagonal em blocos, com blocos diagonais $A_1, \ldots, A_m$, em que*

$$A_j = \lambda_j \in \mathbb{R} \quad ou \quad A_j = \begin{pmatrix} a_j & b_j \\ -b_j & a_j \end{pmatrix},$$

o último caso ocorrendo quando $\lambda_j = a_j + ib_j$ é um autovalor da complexificação $N_{\mathbb{C}} : V_{\mathbb{C}} \to V_{\mathbb{C}}$.

Demonstração: Afirmamos inicialmente que a complexificação $N_{\mathbb{C}}$ de N é uma aplicação normal. De fato, para $u+iv \in V_{\mathbb{C}}$, temos que $N_{\mathbb{C}}^*(u+iv) = N^*u + iN^*v$. Assim,

$$N_{\mathbb{C}}N_{\mathbb{C}}^*(u+iv) = N_{\mathbb{C}}(N^*u+iN^*v) = NN^*u + iNN^*v = N^*Nu + iN^*Nv = N_{\mathbb{C}}^*N_{\mathbb{C}}(u+iv).$$

Se $\lambda \in \mathbb{R}$ é um autovalor de N, então é um autovalor de $N_{\mathbb{C}}$. Considere o polinômio mínimo $m_N(t)$ de N:

$$m_N(t) = (t-\lambda_1)\ldots(t-\lambda_k)(t^2 - 2a_{k+1}t + a_{k+1}^2 + b_{k+1}^2)\ldots(t^2 - 2a_m t + a_m^2 + b_m^2),$$

em que os fatores $(t^2 - 2a_j t + a_j^2 + b_j^2)$ surgem das raízes complexas conjugadas $\lambda_j = a_j \pm ib_j$.

Os termos lineares do polinômio mínimo são diagonalizáveis, como sabemos (veja o teorema 5.5.9).

Consideremos um fator $(t^2 - 2at + a^2 + b^2)$ de segundo grau do polinômio mínimo. Seja $W = \ker(N^2 - 2aN + a^2 + b^2)$. Denotando $S = N|_W$, temos $S : W \to W$ (pelo Teorema da Decomposição Primária). Consideremos então a complexificação $W_{\mathbb{C}}$ de W. De acordo com o Teorema da Decomposição Primária temos

$$\begin{aligned} W_{\mathbb{C}} &= \ker(S_{\mathbb{C}} - \lambda I) \oplus \ker(S_{\mathbb{C}} - \bar{\lambda} I) \\ &= W_\lambda \oplus W_{\bar{\lambda}}. \end{aligned}$$

Seja $\mathcal{C} = \{u_1 + iv_1, \ldots, u_\ell + iv_\ell\}$ uma base ortonormal de W_λ.

Como

$$T(u_k) + iTv_k = T_{\mathbb{C}}(u_k + iv_k) = \lambda(u_k + iv_k) = (a+bi)(u_k + iv_k) = (au_k - bv_k) + i(bu_k + av_k),$$

vemos que

$$Tu_k = au_k - bv_k \quad \text{e} \quad Tv_k = bu_k + av_k.$$

Afirmamos que $\{u_1, v_1, \ldots, u_\ell, v_\ell\}$ é uma base ortogonal de W. A ortogonalidade de $\mathcal{C}$ e dos vetores de $\mathcal{C}$ e de $\overline{\mathcal{C}}$ garante que $u_k + iv_k \perp u_j + iv_j$ para $j \neq k$, e $u_k + iv_k \perp u_j - iv_j$ para todos k, j, com $k, j \in \{1, \ldots \ell\}$.

De fato, temos que

$$0 = \langle u_j + iv_j, u_j - iv_j \rangle = \|u_j\|^2 + i\langle u_j, v_j \rangle + i\langle v_j, u_j \rangle - \|v_j\|^2.$$

A igualdade acima mostra que $\|u_j\| = \|v_j\|$ e que

$$u_j \perp v_j.$$

Da mesma forma,

$$0 = \langle u_j + iv_j, u_k + iv_k \rangle = \langle u_j, u_k \rangle - i\langle u_j, v_k \rangle + i\langle v_j, u_k \rangle + \langle v_j, v_k \rangle,$$

mostrando que

$$\langle u_j, u_k \rangle + \langle v_j, v_k \rangle = 0 \quad \text{e} \quad \langle u_j, v_k \rangle = \langle v_j, u_k \rangle. \tag{7.11}$$

Por outro lado,

$$0 = \langle u_j + iv_j, u_k - iv_k \rangle = \langle u_j, u_k \rangle + i\langle u_j, v_k \rangle + i\langle v_j, u_k \rangle - \langle v_j, v_k \rangle,$$

provando que

$$\langle u_j, u_k \rangle = \langle v_j, v_k \rangle \quad \text{e} \quad \langle u_j, v_k \rangle + \langle v_j, u_k \rangle = 0. \tag{7.12}$$

Decorre de (7.11) e (7.12) que

$$u_j \perp u_k \quad \text{e} \quad v_j \perp v_k.$$

Como $\dim W = 2\ell$ e os vetores $u_k, v_k \in W_{\mathbb{C}}$ e também a W, provamos o afirmado.

□

Aplicamos então os teoremas 7.4.4 e 7.4.5 e obtemos:

Teorema 7.4.6 *Seja $U : V \to V$ uma aplicação unitária definida no espaço euclidiano complexo V. Então:*

(i) Existe uma base ortonormal formada por autovetores de U;

(ii) Os autovalores de U tem valor absoluto igual a 1.

Demonstração: Como $U^*U = I$, U tem inversa $U^{-1} = U^*$. Isso implica que U é normal, possuindo assim uma base ortonormal formada por seus autovetores. Se λ é um autovetor de U associado ao autovalor v, então $\|Uv\| = \|\lambda v\| = |\lambda| \, \|v\|$. Como U é isométrica, $|\lambda| = 1$.

□

Teorema 7.4.7 *Seja $T : V \to V$ um operador ortogonal (isto é, uma isometria) definida no espaço euclidiano real V. Então existe uma base ortonormal $\mathcal{B}$ na qual T é uma matriz diagonal em blocos, com blocos diagonais 1, -1 e blocos 2×2 da forma*

$$\begin{pmatrix} \cos\theta & \operatorname{sen}\theta \\ -\operatorname{sen}\theta & \cos\theta \end{pmatrix}.$$

Demonstração: A demonstração do teorema 7.4.5 mostra a existência de uma base ortonormal com blocos de tamanho 1×1 ou 2×2.

A demonstração do teorema 7.4.6 mostra que os autovalores de T têm valor absoluto igual a 1. Como T é uma isometria, a imagem de uma base ortonormal é uma base ortonormal. Isso mostra que cada coluna das matrizes diagonais 2×2 devem ter norma 1. Mas, de acordo como o teorema 7.4.5, essa matriz tem a forma

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

Como $a^2 + b^2 = 1$, podemos escrever essa matriz na forma

$$\begin{pmatrix} \cos\theta & \operatorname{sen}\theta \\ -\operatorname{sen}\theta & \cos\theta \end{pmatrix}.$$

□

Considerando as definições dadas anteriormente, a matriz A é **anti-auto-adjunta** se $A^* = -A$ (no caso real, **anti-simétrica**) e **normal** se $A^*A = AA^*$.

7.5 O teorema dos valores singulares

Definição 7.5.1 *Uma aplicação linear auto-adjunta* $T : V \to V$ *é* **não-negativa** *se* $\langle Tv, v\rangle \geq 0$ *para todo* $v \in V$*. Nesse caso, escrevemos* $T \geq 0$*. Quando* $\langle Tv, v\rangle > 0$ *para todo* $v \neq 0$*, escrevemos* $T > 0$ *e dizemos que* T *é* **positiva**.

Note que todos os autovalores de uma aplicação positiva (respectivamente, não-negativa) são maiores que zero (resp., maiores ou iguais a zero). De fato, se $T > 0$ e $Tv = \lambda v$, então $\lambda\langle v, v\rangle = \langle Tv, v\rangle > 0$.

Exemplo 7.5.2 Seja $T : V \to W$ uma aplicação linear entre os espaços euclidianos V e W. Como sua adjunta é uma aplicação de W para V, existe a composta $T^*T : V \to V$, que é auto-adjunta e não-negativa. De fato,

$$(T^*T)^* = T^*(T^*)^* = T^*T \quad \text{e} \quad \langle T^*Tv, v\rangle = \langle Tv, Tv\rangle \geq 0.$$

◀

O próximo resultado é, em certo sentido, o análogo à diagonalização de operadores lineares para o caso de aplicações $T : V \to W$ entre espaços euclidianos distintos. Algumas implicações matriciais desse resultado serão estudadas no capítulo 8.

Teorema 7.5.3 (Decomposição em valores singulares)
Sejam V, W *espaços euclidianos e* $T : V \to W$ *uma aplicação linear de posto* r*. Então existem bases ortonormais* $\mathcal{B} = \{v_1, \ldots, v_n\}$ *de* V *e* $\mathcal{C} = \{w_1, \ldots, w_m\}$ *de* W *tais que*

$$\begin{aligned}
Tv_i &= \mu_i w_i \quad \textit{para} \ \ i = 1, \ldots, r, \ \textit{com}\ \mu_i > 0\\
Tv_i &= 0 \quad \textit{para} \ \ i = r+1, \ldots, n\\
T^*w_i &= \mu_i v_i \quad \textit{para} \ \ i = 1, \ldots, r\\
T^*w_i &= 0 \quad \textit{para} \ \ i = r+1, \ldots, m.
\end{aligned}$$

Denotando D_1 *a matriz* $n \times n$

$$D_1 = \begin{pmatrix} \mu_1 & & & & & & \\ & \mu_2 & & & & & \\ & & \ddots & & & & \\ & & & \mu_r & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{pmatrix},$$

a representação $[T]_{\mathcal{B}}^{\mathcal{C}}$ *é, portanto, a matriz* $m \times n$

$$[T]_{\mathcal{B}}^{\mathcal{C}} = \begin{pmatrix} D_1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Demonstração: O exemplo 7.5.2 mostra que $T^*T : V \to V$ é uma aplicação não-negativa. Temos $\ker T = \ker(T^*T)$. De fato,

$$Tv = 0 \Leftrightarrow \langle Tv, Tu\rangle = 0\ \forall\ u \in V \Leftrightarrow \langle T^*Tv, u\rangle\ \forall\ u \in V \Leftrightarrow T^*Tv = 0.$$

Isso mostra que posto$(T^*T) = n - \dim\ker T^*T = n - \dim\ker T = r$.

Uma vez que T^*T é um operador normal, o teorema 7.4.4 garante que existe uma base ortonormal $\mathcal{B} = \{v_1, \ldots, v_n\}$ de V tal que

$$T^*T(v_i) = \mu_i^2 v_i,\ i = 1, \ldots, r \quad \text{e} \quad T^*T(v_i) = 0,\ i = r+1, \ldots, n.$$

(Note que os autovalores do operador T^*T são não-negativos, como vimos).

Defina, para $i = 1, \ldots, r$,

$$w_i = \frac{1}{\mu_i}T(v_i).$$

Então $Tv_i = \mu_i w_i$. (Note que, se $i \in \{r+1, \ldots, n\}$, como $0 = \langle T^*Tv_i, v_i\rangle = \langle Tv_i, Tv_i\rangle$, temos $Tv_i = 0$. Assim, os vetores $v_{r+1}, \ldots, v_n$ formam uma base ortonormal de $\ker T = \ker T^*T$, se esse subespaço é não-vazio).

Mostraremos que $\{w_1, \ldots, w_r\}$ é base ortonormal de $\mathcal{I}m\ T$. Claramente esses vetores pertencem à imagem de T e, se $i, j \in \{1, \ldots, r\}$, então

$$\langle w_i, w_j\rangle = \frac{1}{\mu_i\overline{\mu_j}}\langle Tv_i, Tv_j\rangle = \frac{1}{\mu_i\overline{\mu_j}}\langle T^*Tv_i, v_j\rangle = \frac{1}{\mu_i\overline{\mu_j}}\langle \mu_i^2 v_i, v_j\rangle = \frac{\mu_i}{\overline{\mu_j}}\langle v_i, v_j\rangle = \frac{\mu_i}{\overline{\mu_j}}\delta_{ij}.$$

Como $\dim\mathcal{I}m\ T = r$, provamos o afirmado.

Além disso,

$$T^*w_i = T^*\left(\frac{1}{\mu_i}Tv_i\right) = \frac{1}{\mu_i}T^*Tv_i = \mu_i v_i.$$

Seja $\{w_{r+1}, \ldots, w_m\}$ uma base ortonormal de $\ker T^*$. Então $T^*w_i = 0$ para $i = r+1, \ldots, n$. Uma vez que $\ker T^* = (\mathcal{I}m\ T)^\perp$, os vetores $\{w_1, \ldots, w_m\}$ formam uma base ortonormal de W, de acordo com o teorema 6.4.2. Isso completa a prova.

□

Usualmente a base $\mathcal{B}$ é ordenada de modo que $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_r$. Veja um exemplo na seção 8.5.

7.6 Exercícios

1. Sejam $\mathcal{B} = \{x_1, \ldots, x_n\}$ e $\mathcal{C} = \{y_1, \ldots, y_n\}$ duas bases no espaço vetorial V. Se a matriz de B na base $\mathcal{B}$ é A, mostre que a matriz de B na base $\mathcal{C}$ é P^TAP, em que P é a matriz mudança da base $\mathcal{B}$ para a base $\mathcal{C}$. Isso significa que formas bilineares mudam suas representações em bases de maneira diferente da mudança de base de matrizes!

2. Seja V um espaço euclidiano real. Mostre que, dada uma forma quadrática $q : V \to \mathbb{R}$, então

$$B(u, v) = \frac{1}{2}[q(u + v) - q(u) - q(v)] \tag{7.13}$$

é a (única) forma bilinear **simétrica** $B : V \times V \to \mathbb{R}$ tal que $q(v) = B(v, v)$. A identidade (7.13) é a **identidade de polarização**.

3. Dada a forma quadrática $ax^2 + bxy + cy^2$, encontre a matriz que a representa.

4. Seja V um espaço vetorial complexo. Além das formas bilineares definidas em V, são importantes as formas $B : V \times V \to \mathbb{C}$ tais que para quaisquer $\alpha \in \mathbb{C}$ e $u_1, u_2, v_1, v_2 \in V$,

   (i) $B(\alpha u_1 + u_2, v) = \alpha B(u_1, v) + B(u_2, v)$;

   (ii) $B(u, \alpha v_1 + v_2) = \overline{\alpha} B(u, v_1) + B(u, v_2)$.

   Essas são as formas **sesquilineares** definidas em V. Denotaremos $\mathcal{B}(V)$ o conjunto das formas sesquilineares em V.

   Uma forma sesquilinear é **hermitiana** se $B(u, v) = \overline{B(v, u)}$ para quaisquer $u, v \in V$. (Um produto interno definido no espaço euclidiano V é um exemplo de uma forma sesquilinear hermitiana).

   Verifique:

   (a) Seja V um espaço euclidiano. Mostre que existe um único operador linear $T : V \to V$ tal que
   $$B(u, v) = \langle Tu, v\rangle$$
   e a aplicação $B \mapsto T$ é um isomorfismo entre $\mathcal{B}(V)$ e $\mathcal{L}(V)$. (Em particular, escolhida uma base de V, a forma sesquilinear B é representada por uma matriz A).

   (b) Se B é uma forma (sesquilinear) hermitiana, definindo $q(v) = B(v, v)$ (chamada **forma quadrática hermitiana**), vale
   $$B(u, v) = \frac{1}{4}[q(u + v) - q(u - v)] + \frac{i}{4}[q(u + iv) - q(u - iv)], \tag{7.14}$$
   chamada **identidade de polarização** (veja os exercícios 8 do capítulo 6 e 2, desse capítulo);

   (c) Seja $B \in \mathcal{B}(V)$. Então B é hermitiana se, e somente se, $B(u, u) \in \mathbb{R}$ para todo $u \in V$. Se V é um espaço euclidiano, conclua que $A : V \to V$ é auto-adjunto se, e somente se, $\langle Tu, u\rangle \in \mathbb{R}$ para todo $u \in \mathbb{R}$;

   (d) O Teorema de Sylvester 7.2.1 (incluindo a Lei de Inércia) pode ser generalizado, se $B(u, v)$ é uma forma sesquilinear hermitiana;

   (e) Generalize, tendo em vista o item (a), o teorema 7.2.2: mostre que se A é uma matriz hermitiana A, então existe uma matriz invertível $M \in \mathbb{M}_{n\times n}(\mathbb{C})$ tal que $M^*AM = D$, sendo D uma matriz diagonal.

   (Note que, se V é um espaço vetorial real, toda forma sesquilinear é bilinear e uma forma hermitiana é simétrica. Em particular, os resultados desse exercício são válidos para espaços vetoriais reais).

5. Seja $H : V \to V$ uma matriz auto-adjunta definida no espaço complexo V. Tendo em vista o exercício 4, verifique que o teorema 7.3.3 pode ser generalizado: existe uma matriz unitária M tal que $M^*HM = D$, sendo D a matriz diagonal cujos elementos diagonais são os autovalores de H.

6. Defina a equivalência de duas matrizes A e B, denotado $A \cong B$, se existe uma matriz invertível M tal que $A = M^*BM$. Mostre que assim está definida uma relação de equivalência em $\mathbb{M}_{n\times n}(\mathbb{C})$.

7. O teorema 7.2.2 mostra que toda matriz simétrica é equivalente a uma matriz diagonal. Dada a equivalência entre os teoremas 7.2.1 e 7.2.2, podemos concluir que a Lei de Inércia é uma afirmação sobre matrizes simétricas. Ela garante que, no teorema 7.2.2, o número de termos positivos, negativos e nulos na matriz diagonal D independe da mudança de variável utilizada. Por outro lado, sabemos que se D é a diagonalização da matriz A, então os elementos diagonais de D são os autovalores de A. Mas sabemos que os autovalores de A independem da base na qual a matriz é representada. Isso não implica a Lei de Inércia?

8. Explique a relação $M^*AM = D$ em termos de mudanças de coordenadas.

9. Considere a matriz simétrica
$$A = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 4 \end{pmatrix}.$$
Ache uma matriz ortogonal (isto é, $P^T = P^{-1}$) e uma matriz diagonal D tais que
$$P^{-1}AP = D.$$

10. Seja $T : V \to W$ uma aplicação linear entre espaços euclidianos.
    (a) Mostre que se T é injetiva, então T^*T possui inversa;
    (b) Mostre que $\mathcal{I}m\, T^* = \mathcal{I}m\,(T^*T)$;
    (c) Mostre que se T é sobrejetiva, então TT^* possui inversa.

11. Sejam $S, T : V \to V$ aplicações lineares, S sendo normal. Suponha que $ST = TS$. Mostre que S e T são simultaneamente diagonalizáveis por uma (mesma) matriz mudança de base ortonormal.

12. Seja $T : V \to V$ uma aplicação linear no espaço euclidiano V. Mostre que
$$\|T\| = \sup_{\|x\|=1=\|y\|} |\langle Tx, y\rangle|.$$
Conclua então que $\|T\| = \|T^*\|$.

13. Mostre que se $T : V \to V$ possui uma base ortonormal $v_1, \ldots, v_n$ constituída de autovetores, então T é normal.

14. Seja $T : V \to V$ um operador linear definido no espaço real V. Mostre que existe uma base ortonormal $\mathcal{B}$ na qual $T_{\mathcal{B}}$ é diagonal se, e somente se, T é auto-adjunto.

15. No decorrer da demonstração do teorema de Sylvester 7.2.1 se introduziu o conceito de forma quadrática positiva no espaço V: a forma quadrática é positiva se $q(v) > 0$ para todo $0 \neq v \in V$. Como sabemos, à forma quadrática q está associada a matriz auto-adjunta A, definida por $q(v) = \langle v, Av\rangle$. O conceito de uma aplicação auto-adjunta positiva foi definido na definição 7.5.1. Mostre que essas duas definições coincidem no caso da matriz A.

16. Seja $T : V \to V$ uma aplicação auto-adjunta e $\{v_1, \ldots, v_n\}$ uma base ortonormal formada por autovetores de T. Mostre que se todos os autovalores de T são positivos (resp., não-negativos), então $T > 0$ (resp., $T \geq 0$).

17. Se $T : V \to V$ satisfaz $T \geq 0$ e $\langle Tv, v\rangle = 0$, mostre que $Tv = 0$.

18. Mostre que um operador T é positivo se, e somente se, $T \geq 0$ e T é invertível.

19. Uma aplicação linear $S : V \to V$ é uma **raiz quadrada** da aplicação linear $T : V \to V$ quando $S^2 = T$. Mostre que toda aplicação linear auto-adjunta $T : V \to V$ que é não-negativa possui uma única raiz quadrada não-negativa. Essa é positiva se, e somente se T é positiva.

20. Seja $T : V \to V$ uma aplicação linear invertível. Mostre que existe uma única aplicação auto-adjunta positiva (isto é, $\langle Pv, v\rangle > 0$ para todo $v \neq 0$) e um operador unitário (ortogonal, se V é um espaço real) tais que

$$T = PU.$$

Essa é a **decomposição polar** de T.

21. Qual a relação entre os autovalores de T^*T e os de TT^*?

Capítulo 8

Decomposições Matriciais

Nesse capítulo estudaremos algumas decomposições matriciais. Os resultados que apresentaremos são bastante úteis na Álgebra Linear Numérica.

8.1 O método de Gauss

A primeira sub-seção revê a teoria básica de sistemas lineares.

8.1.1 Sistemas lineares e escalonamento

Para $1 \leq i \leq m$ e $1 \leq j \leq n$, suponhamos conhecidos os valores a_{ij} e os valores b_j. Um **sistema linear** em m equações e n incógnitas procura a solução $x_1, \ldots, x_n$ que satisfaz

$$\begin{array}{ccccccc}
a_{11}x_1 & + & \ldots & + & a_{1n}x_n & = & b_1 \\
a_{21}x_1 & + & \ldots & + & a_{2n}x_n & = & b_2 \\
 & & \vdots & & & = & \vdots \\
a_{m1}x_1 & + & \ldots & + & a_{mn}x_n & = & b_m.
\end{array}$$

Em termos de matrizes, esse sistema pode ser escrito como

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix},$$

ou,

$$Ax = b$$

Mais sinteticamente ainda, podemos representar esse sistema por uma única matriz, chamada **matriz aumentada do sistema**:

$$\mathcal{A} = (A \mid b) = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right)$$

É fácil verificar que as seguintes operações sobre as linhas da matriz $\mathcal{A}$ não alteram o conjunto de soluções[1] do sistema $Ax = b$:

(i) Transpor as linhas i e j;

(ii) Multiplicar a linha i por um escalar não-nulo;

(iii) Substituir a linha j por sua soma com um múltiplo da linha i.

As operações $(i) - (ii) - (iii)$ são as **operações elementares** sobre as linhas da matriz $\mathcal{A}$.

Uma sucessiva aplicação de operações elementares sobre a matriz $\mathcal{A}$ pode fazer com que essa matriz se transforme numa matriz com as seguintes propriedades:

(i) se o primeiro elemento não-nulo da linha i (chamado pivô da linha i) ocorre na coluna j, então o primeiro elemento da linha $i + \ell$ ocorre numa coluna $k > j$, para todo $\ell \in \mathbb{N}^*$;

(ii) o pivô de cada linha é igual a 1.

De fato, se existe algum elemento não nulo na primeira coluna de $\mathcal{A}$, ao aplicarmos as operações elementares (i) e (ii) obtemos uma nova matriz $\mathcal{A}' = (a'_{ij})$, com $a'_{11} = 1$. A aplicação da operação elementar (iii) torna possível transformar em zero qualquer outro elemento não-nulo da primeira coluna. O resultado então segue por indução sobre o número de linhas de $\mathcal{A}$. Essa forma da matriz $\mathcal{A}$ é chamada forma **escalonada** e a sucessão de operações elementares utilizadas é um **escalonamento** da matriz A.

Suponhamos agora que a matriz $\mathcal{A}$ esteja na **forma escalonada**. Se o pivô for o único elemento não-nulo de cada coluna, dizemos que a matriz está na forma **escalonada reduzida por linhas**. Aplicando as operações elementares (i) e (iii), podemos fazer com que uma matriz na forma escalonada atinja sua forma reduzida por linhas. De fato, consideremos o pivô da segunda linha de $\mathcal{A}$. A aplicação da operação elementar (iii) torna possível zerar o elemento que está acima do pivô, mantendo ainda a matriz na forma escalonada. A demonstração agora segue por indução.

Dois sistemas $Ax = b$ e $A'x = b'$ são **equivalentes** se eles possuem as mesmas soluções. Se as formas escalonadas reduzidas por linhas de ambos os sistemas possuem linhas da forma

$$(0 \mid \beta), \tag{8.1}$$

ambos não possuem soluções e são, portanto, equivalentes.

Suponhamos agora que, na forma escalonada reduzida por linhas $(R \mid c)$ do sistema $Ax = b$, a equação (8.1) não se verifique. Afirmamos que o sistema possui solução. Na matriz R, o número

[1] Note que $x = (x_1, x_2, \ldots, x_n)$ satisfaz

$$\begin{array}{ccccccc} a_{i1}x_1 & + & \ldots & + & a_{in}x_n & = & b_i \\ a_{j1}x_1 & + & \ldots & + & a_{jn}x_n & = & b_j \end{array}$$

se, e somente se, satisfaz

$$\begin{array}{ccccccc} a_{i1}x_1 & + & \ldots & + & a_{in}x_n & = & b_i \\ (a_{j1} + \alpha a_{i1})x_1 & + & \ldots & + & (a_{jn} + \alpha a_{in})x_n & = & b_j + \alpha b_i. \end{array}$$

de pivôs (isto é, de linhas não-nulas) é, no máximo, igual ao número de colunas de R. Se o número de pivôs é igual ao número de colunas da matriz R, então a matriz $(R \mid c)$ é da forma

$$\left(\begin{array}{c|c} I & c' \\ 0 & 0 \end{array}\right),$$

em que I é matriz identidade $n \times n$ e 0 a matriz nula de ordem $(m-n) \times n$. Assim, o sistema $Ax = b$ é equivalente ao sistema $x = c'$ e apenas uma solução. Se, por outro lado, o número de pivôs é menor do que o número de colunas da matriz R, o sistema possui infinitas soluções. De fato, suponhamos que o sistema $(R \mid c)$ tenha um pivô a menos. Desprezando as linhas nulas desse sistema, ele tem a forma

$$\begin{array}{r} \\ \\ \\ \text{linha } i \rightarrow \\ \\ \\ \\ \end{array} \begin{pmatrix} 1 & 0 & \dots & 0 & \alpha_1 & 0 & \dots & 0 \\ & 1 & \dots & 0 & \alpha_2 & 0 & \dots & 0 \\ & & & \vdots & \vdots & & & \vdots \\ & & & 1 & \alpha_i & 0 & \dots & 0 \\ & & & & 0 & 1 & \dots & 0 \\ & & & & & & & \vdots \\ & & & & & & & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ x_{i+1} \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} c'_1 \\ c'_2 \\ \vdots \\ c'_i \\ c'_{i+2} \\ \vdots \\ c'_n \end{pmatrix}.$$

$$\begin{array}{c} \uparrow \\ \text{coluna } i+1 \end{array}$$

Assim, sua solução é

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ x_{i+1} \\ x_{i+2} \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} c'_1 \\ c'_2 \\ \vdots \\ c'_i \\ 0 \\ c'_{i+2} \\ \vdots \\ c'_{n-1} \end{pmatrix} - x_{i+1} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_i \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} =: d - x_{i+1}\alpha. \tag{8.2}$$

Em outras palavras, a perda de um pivô gerou um grau de liberdade para o sistema (o valor de x_{i+1}) e, com isso, uma infinidade de soluções, dadas como a soma dos vetores na expressão (8.2). Essa expressão mostra, em particular, a unicidade da forma escalonada reduzida por linhas de qualquer sistema que perca um pivô: se um dos valores α_j for diferente, o vetor α será diferente. Por indução, se o sistema perder k pivôs, a sua solução será a soma de um vetor fixo com uma combinação linear arbitrária de k vetores (tendo, portanto, k graus de liberdade). Essa expressão acarreta, em particular, a unicidade da forma escalonada reduzida por linhas de um sistema que perde k pivôs.

Quer dizer, dois sistemas lineares são equivalentes ou se ambos não possuem soluções ou se, eliminadas as linhas nulas existentes, eles possuem a mesma forma escalonada reduzida por linhas.

8.1.2 Matrizes elementares e a decomposição LU

Consideremos agora uma matriz $m \times n$ A. Vamos mostrar que os resultados da subseção anterior podem ser interpretados como uma decomposição da matriz A.

Uma matriz é **elementar** se ela pode ser obtida da matriz identidade $m \times m$ através da aplicação de uma operação elementar.

Proposição 8.1.1 *Seja e um operação elementar sobre (as linhas da) a matriz A (de ordem $m \times n$) e E a matriz elementar $e(I)$, sendo I a matriz identidade $m \times m$.*

Demonstração: A demonstração deve ser feita com relação a cada uma das operações elementares. Faremos isso apenas no caso da operação elementar (iii): a linha j será substituída pela soma da linha j com c vezes a linha i.

A matriz E, nesse caso, é dada por

$$E = \begin{array}{cc} \begin{pmatrix} 1 & 0 & & \dots & & & 0 \\ \vdots & & & & & & \vdots \\ 0 & \dots & c & \dots & 1 & \dots & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & & \dots & & & 1 \end{pmatrix} & \begin{matrix} \\ \\ \leftarrow \text{linha } j \\ \\ \\ \end{matrix} \\ \begin{matrix} \uparrow \\ \text{coluna } j \end{matrix} & \end{array}$$

Então

$$\begin{aligned} EA &= \begin{pmatrix} 1 & 0 & & \dots & & & 0 \\ \vdots & & & & & & \vdots \\ 0 & \dots & c & \dots & 1 & \dots & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & & \dots & & & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & & & \vdots \\ a_{j1} & a_{j2} & \dots & a_{jn} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \\ &= \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & & & \vdots \\ a_{j1} + ca_{i1} & a_{j2} + ca_{i2} & \dots & a_{jn} + ca_{in} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, \end{aligned}$$

que é justamente $e(A)$.

□

Suponhamos que E seja uma matriz elementar obtida por meio da operação elementar (ii) ou (iii). É fácil verificar que tanto a matriz E como sua inversa (que existe!) são matrizes triangulares inferiores.

Tendo em vista a proposição 8.1.1, dada uma matriz $m \times n$ A, obtemos a forma escalonada da matriz A ao multiplicá-la por matrizes elementares $E_k E_{k-1} \dots E_2 E_1$. Quer dizer,

$$(E_k E_{k-1} \dots E_2 E_1) A = U,$$

em que $U = (u_{ij})$ tem todos os seus elementos abaixo da diagonal u_{ii} iguais a zero.

Suponhamos que, nesse processo de levar a matriz A a sua forma escalonada, a operação elementar (i) não tenha sido utilizada. Uma vez que a matriz $E_kE_{k-1}\ldots E_2E_1$ tem inversa e sua inversa é uma matriz triangular inferior (veja exercícios 2 e 3), obtemos que

$$A = LU$$

em que a matriz L é triangular inferior e a matriz U é "triangular superior". Essa é a decomposição **LU** da matriz A.

Observação 8.1.2 A decomposição $A = LU$ é usualmente feita para matrizes quadradas A. Nesse caso, a matriz U é uma autêntica matriz triangular superior.

Se A é uma matriz $m \times n$ e se no seu escalonamento não foi utilizada a operação elementar (i), a decomposição LU pode ser atingida unicamente por meio da operação elementar (iii): não há necessidade de transformar em 1 o primeiro elemento não nulo de cada linha. Assim, suponhamos que por meio das matrizes elementares E_1,...,E_k todos os elementos abaixo do pivô de cada linha tenham sido anulados até a coluna $j-1$, e que o pivô da coluna j esteja na linha i, com $i \le j$. Se $\ell > i$, para anularmos o elemento $b_{\ell j}$ da matriz $(b_{ij}) = E_k \ldots E_1 A$, substituímos a linha ℓ pela linha ℓ somada a $-c$ vezes a linha i. A essa operação corresponde corresponde a matriz elementar

$$\begin{array}{r c}
 & \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & \vdots & \ddots & & \\ & & -c_{\ell,j} & \cdots & 1 & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \\
\end{array}$$

$$\begin{array}{l}
\text{linha } i \rightarrow \\
\text{linha } \ell \rightarrow
\end{array}
\qquad
\begin{array}{c}
\uparrow \\
\text{coluna } j
\end{array}$$

O valor de $c_{\ell,j}$ é $b_{\ell j}/b_{ij}$, se $b_{\ell j}$ e b_{ij} são os primeiros elementos não-nulos das linhas ℓ e i, respectivamente, da matriz $E_k \ldots E_1 A$. Se multiplicarmos todas as matrizes que anulam os elementos $b_{\ell j}$, com $\ell > i$, obteremos a matriz

$$Q_j = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & -c_{i+1,j} & & \\ & & \vdots & \ddots & \\ & & -c_{i+r,j} & & 1 \end{pmatrix}.$$

É fácil verificar que $L_j = Q_j^{-1}$ existe e tem o mesmo formato da matriz acima. Decorre daí que, na decomposição LU da matriz A, todos os elementos da diagonal principal da matriz L são iguais a 1. ◀

Seja A uma matriz $m\times n$. Suponhamos que, no escalonamento de A, tenhamos exatamente n pivôs e que não tenhamos utilizado a operação elementar (i). Isso implica que, na decomposição LU da matriz A, os elementos diagonais da matriz $m \times m$ L são todos iguais a 1, enquanto os elementos diagonais da matriz U são justamente os pivôs. Podemos então escrever a matriz A numa forma mais simétrica: se

$$U = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & u_{nn} \\ 0 & 0 & \cdots & 0 \\ \vdots & & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix},$$

com $u_{ii} \neq 0$, então podemos decompor

$$U = DU' = \begin{pmatrix} u_{11} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & u_{22} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & & u_{nn} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} 1 & u_{12}/u_{11} & \cdots & u_{1n}/u_{11} \\ 0 & 1 & \cdots & u_{2n}/u_{22} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \\ \vdots & & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix},$$

em que D é uma matriz $m \times m$ e U' uma matriz $m \times n$, com elementos "diagonais" iguais a 1. Temos, assim,

$$A = LDU'.$$

É usual escrever $A = LDU$, chamada **decomposição LDU** da matriz A.

Proposição 8.1.3 *Seja A uma matriz $m \times n$. Se $A = LU$ e $A = \overline{L}\overline{U}$, com $L, \overline{L}$ matrizes $m \times m$ triangulares inferiores com elementos diagonais iguais a 1 e $U, \overline{U}$ matrizes triangulares superiores com elementos "diagonais" não nulos. Então $L = L'$ e $U = U'$. Em particular, a decomposição LDU de uma matriz é única.*

Demonstração: Como a matriz L possui inversa, temos $U = (L^{-1}\overline{L})\overline{U}$. A matriz quadrada $L^{-1}\overline{L}$ é triangular inferior e tem elementos diagonais iguais a 1. Vamos mostrar que $L^{-1}\overline{L} =: R = (r_{ij})$ é a matriz identidade. Temos $r_{i1} = 0$ se $i \neq 1$, o que pode ser comprovado multiplicando a linha i de R pela primeira coluna de $\overline{U}$, pois $R\overline{U}$ é uma matriz triangular inferior e $\bar{u}_{11} \neq 0$. Da mesma forma, multiplicando as linha de R pela segunda coluna de $\overline{U}$, verificamos que $r_{i2} = 0$ se $i \neq 2$ e assim sucessivamente. Logo $R = I$ e $U = \overline{U}$.

Daí decorre, em particular, que os elementos diagonais de U e U' são iguais. Se $D = (d_{ij})$ é a matriz diagonal $m \times m$ com $d_{ii} = u_{ii}$ para $i = 1, \ldots, n$ e $d_{jj} = 0$ se $j > n$, então podemos escrever $U = D\overline{U}$, as matrizes D e $\overline{U}$ tendo o formato dado na decomposição LDU da matriz A.

□

Definição 8.1.4 *Seja* $A = (a_{ij})$ *uma matriz* $m \times n$. *Uma* **sub-matriz** *de* A *é a matriz obtida de* A *ao se omitir algumas de suas linhas e/ou colunas. Denotaremos* A_r *a matriz* (a_{ij}), $1 \leq i, j \leq r \leq n$. *A sub-matriz* A_r *é a* **sub-matriz principal** *de* A *de ordem* r.

Proposição 8.1.5 *Seja* A *uma matriz* $m \times n$ *tal que todas suas sub-matrizes principais* A_r *sejam invertíveis. Então* A *tem uma decomposição* LU.

Demonstração: Como $a_{11} = A_1$, o elemento a_{11} é o pivô da primeira linha. Existe então uma matriz invertível E, obtida ao se aplicar sucessivamente a operação elementar (iii) de modo a anular todos os elementos de A abaixo do pivô. Temos então que

$$EA = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & b_{22} & \cdots & b_{2n} \\ \vdots & & & \vdots \\ 0 & b_{m2} & \cdots & b_{mn} \end{pmatrix}.$$

Claramente a sub-matriz principal de EA

$$\begin{pmatrix} a_{11} & a_{12} \\ 0 & b_{22} \end{pmatrix}$$

resulta da sub-matriz principal de A

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

mediante a aplicação de uma operação elementar do tipo (iii). Em particular, aquela sub-matriz principal de EA é invertível, pois a sub-matriz de A é invertível (por hipótese). Daí decorre que $b_{22} \neq 0$, mostrando que b_{22} é um pivô da segunda linha de EA. A prova agora segue por indução. $\square$

Suponhamos agora que, ao levarmos a matriz A a sua forma escalonada seja necessária a aplicação da operação elementar (i). Então não é possível decompor a matriz A na forma LU. Entretanto, podemos considerar as matrizes elementares que fazem as transposições de linhas necessárias para o escalonamento da matriz A. Cada matriz dessas é ortogonal e se consideramos a matriz P, multiplicação de todas essas matrizes, obtemos uma matriz, chamada **matriz de permutação**.

Consideremos então a matriz PA. Com essa permutação das linhas de A, é possível levar a matriz A a uma forma triangular superior por meio unicamente da operação elementar (iii). Assim, para a matriz PA vale:

$$PA = LU.$$

Como a matriz P é ortogonal, temos então

$$A = P^T LU.$$

8.2 A decomposição de Cholesky

Como vimos na seção 7.5, uma aplicação auto-adjunta é positiva se todos os seus autovalores são positivos.

Seja A uma matriz (real) simétrica, representação matricial de um operador auto-adjunto positivo $T : \mathbb{R}^n \to \mathbb{R}^n$ (resp., não-negativo) com relação à base canônica do $\mathbb{R}^n$. Como $D = P^T AP$ para alguma matriz ortogonal A e uma matriz diagonal D, temos que $\det A > 0$ (resp., $\det A \geq 0$). Uma matriz A simétrica com todos os autovalores positivos é chamada matriz **positiva-definida**.

Observação 8.2.1 Estamos denominando a matriz A de positiva-definida, pois o termo matriz positiva é utilizado em outra situação (geralmente associada ao teorema de Perron). ◀

Lema 8.2.2 *Seja A uma matriz $n \times n$ simétrica positiva-definida. Então as sub-matrizes principais A_r são positivas-definidas (e, portanto, $\det A_r > 0$) para $1 \leq r \leq n$.*

Demonstração: Seja $x_r = (x_1, \ldots, x_r) \in \mathbb{R}^r$ um vetor não-nulo arbitrário e defina $x = (x_1, \ldots, x_r, 0, \ldots, 0) \in \mathbb{R}^n$. Como

$$\langle x_r, A_r x_r \rangle = \langle x, Ax \rangle$$

e A é positiva-definida, o resultado segue. □

Note que o lema 8.2.2 combinado com a proposição 8.1.5 garante que uma matriz positiva-definida A possui decomposição LU, obtida unicamente mediante a sucessiva aplicação da operação elementar (iii) à matriz A. Em particular, A possui uma fatoração LDU, a matriz diagonal $D = (d_{ii})$ tendo seus elementos diagonais positivos. Mas, como a matriz A é simétrica, temos

$$LDU = A = A^T = U^T A L^T.$$

Pela proposição 8.1.3 temos $L^T = U$, de modo que $A = LDL^T$. Definindo $D^{1/2}$ como a matriz

$$D^{1/2} = \begin{pmatrix} \sqrt{d_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{d_{22}} & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & \sqrt{d_{nn}} \end{pmatrix}.$$

Mas então $A = LDL^T = (LD^{1/2})(D^{1/2}L^T) = L_1 L_2$, a matriz L_1 sendo triangular inferior e a matriz L_2 sendo triangular superior. Como $A = A^T$, segue que $L_2 = L_1^T$, mostrando que

$$A = LL^T,$$

chamada **decomposição de Cholesky** da matriz A.

Assim, uma matriz $n \times n$ positiva-definida tem duas decomposições: a decomposição $A = LDU$ e a decomposição de Cholesky $A = L_1 L_1^T$. Já vimos que $L_1 = LD^{1/2}$, o que nos mostra como obter a decomposição de Cholesky da matriz A.

O próximo resultado caracteriza as matrizes positivas-definidas e apresenta um resumo dos resultados obtidos nessa seção:

Proposição 8.2.3 *Seja A uma matriz simétrica $n \times n$. As seguintes afirmações são equivalentes:*

(i) *A é positiva-definida;*

(ii) *As sub-matrizes principais $A_1, \ldots, A_n$ têm determinante positivo;*

(iii) *A matriz A tem uma decomposição LDU, com os elementos diagonais da matriz diagonal D todos positivos;*

(iv) *A tem uma decomposição de Cholesky $A = LL^T$, sendo L uma matriz triangular inferior com elementos diagonais positivos.*

Demonstração: Já vimos as implicações $(i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (iv)$.

Seja agora $x \in \mathbb{R}^n$ um vetor arbitrário e $y = L^T x$. Como a matriz L^T possui inversa, $y \neq 0$. Assim

$$\langle x, Ax \rangle = x^T (LL^T x) = (x^T L)(L^T x) = y^T y = \|y\|^2 > 0.$$

Isso mostra que $(iv) \Rightarrow (i)$.

□

8.3 A decomposição de Schur

Seja A uma matriz $n \times n$ no corpo $\mathbb{C}$.

Teorema 8.3.1 (Schur)
Existe uma matriz unitária U tal que $T = U^*AU$ é uma matriz triangular superior.

Demonstração: Faremos indução em n, o resultado sendo óbvio para $n = 1$. Suponhamos válido para uma matriz $k \times k$ qualquer e consideremos A, matriz $(k+1) \times (k+1)$. Seja w_1 um autovetor unitário associado ao autovalor λ_1 de A. O processo de ortogonalização de Gram-Schmidt assegura a existência de uma base ortonormal $\{w_1, w_2, \ldots, w_{k+1}\}$ para $\mathbb{C}^{k+1}$. A matriz R, cuja i-ésima coluna é o vetor w_i, é unitária (veja o exercício 20 do capítulo 6). Consideremos então $R^*AR = (R^*A)R$. A primeira coluna dessa matriz é R^*Aw_1. Mas $R^*Aw_1 = \lambda_1 R^* w_1 = \lambda_1 e_1$, pois as linhas de R^* são dadas pelos vetores $\overline{w_1}, \ldots, \overline{w_{k+1}}$. Assim, a matriz R^*AR tem a forma

$$\begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & & & \\ \vdots & & S & \\ 0 & & & \end{pmatrix},$$

em que S é uma matriz $k \times k$. Pela hipótese de indução, existe uma matriz unitária V_1 tal que $T_1 = V_1^* S V_1$ é uma matriz triangular superior. Definimos então

$$V = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & V_1 & \\ 0 & & & \end{pmatrix}.$$

Claramente V é unitária e

$$\begin{aligned} V^*(R^*AR)V &= \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & V_1^* & \\ 0 & & & \end{pmatrix} \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & & & \\ \vdots & & S & \\ 0 & & & \end{pmatrix} \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & V_1 & \\ 0 & & & \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & & & \\ \vdots & & V_1^*SV_1 & \\ 0 & & & \end{pmatrix} = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & & & \\ \vdots & & T_1 & \\ 0 & & & \end{pmatrix} = T, \end{aligned}$$

uma matriz triangular superior. Definimos então $U = RV$. A matriz U é unitária, pois

$$U^*U = (RV)^*(RV) = V^*R^*RV = I.$$

Isso completa a demonstração.

□

A demonstração apresentada continua válida se A é uma matriz real cujos autovalores estão no corpo $\mathbb{R}$. Uma prova alternativa do teorema de Schur é indicada no exercício 6. Note que o teorema pode também ser formulado para aplicações lineares ao invés de matrizes.

Corolário 8.3.2 *Se A é uma matriz auto-adjunta, então existe uma matriz unitária U tal que $U^*AU = D$, sendo D uma matriz diagonal. Se A é uma matriz real, a matriz U é ortogonal.*

Demonstração: De acordo com o teorema de Schur 8.3.1, existe uma matriz unitária U tal que $U^*AU = T$, sendo T uma matriz triangular superior. Mas

$$T^* = (U^*AU)^* = U^*A^*U = U^*AU = T,$$

de acordo com a proposição 6.5.3. Isso mostra que T é auto-adjunta e, portanto, uma matriz diagonal.

Se A é real, todos os autovalores de A são reais e, portanto, também seus autovetores. Isso implica que a matriz U é ortogonal.

□

8.4 A decomposição QR

O processo de ortogonalização de Gram-Schmidt pode ser interpretado como uma decomposição de uma matriz cujas colunas são linearmente independentes.

Teorema 8.4.1 (A decomposição QR)
Seja A uma matriz $m \times n$ de posto n. Então

$$A = QR,$$

em que Q é uma matriz $m \times n$ com colunas ortonormais e R é uma matriz $n \times n$ triangular superior com elementos diagonais positivos.

Demonstração: Sejam $v_1, \ldots, v_n$ as colunas da matriz A. Como essa matriz tem posto n, esses vetores são linearmente independentes em $\mathbb{K}^m$. Aplicando o processo de ortogonalização de Gram-Schmidt 6.3.4 a esses vetores, obtemos os vetores ortonormais $q_1, \ldots, q_n \in \mathbb{K}^m$, dados por

$$q_k = r_{kk}\left(v_k - \sum_{i=1}^{k-1} r_{ik}q_i\right), \quad (k = 1, \ldots, n)$$

em que $r_{ik} = \langle v_k, q_i \rangle$ para $i = 1, \ldots, k-1$ e $(1/r_{kk})$ é a norma do vetor $v_k - \sum_{i=1}^{k-1} r_{ik}q_i$. Mas isso quer dizer que

$$\begin{aligned} v_1 &= r_{11}q_1 \\ v_2 &= r_{12}q_1 + r_{22}q_2 \\ \vdots &= \qquad \vdots \\ v_n &= r_{1n}q_1 + \ldots + r_{nn}q_n. \end{aligned}$$

Definindo Q como a matriz cujas colunas são os vetores $q_1, \ldots, q_n$ e R a matriz triangular superior

$$R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{21} & \cdots & r_{2n} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & r_{nn} \end{pmatrix} = (r_1 \ r_2 \ \cdots \ r_n),$$

temos que a j-ésima coluna da matriz QR é

$$QRe_j = Qr_j = r_{1j}q_1 + r_{2j}q_2 + \ldots + r_{jj}q_j = v_j.$$

Isso mostra que $QR = A$, completando a demonstração.

□

8.5 A decomposição em valores singulares

Seja A é uma matriz $m \times n$. A determinação do posto de A, através do escalonamento dessa muitas vezes não é viável numericamente, devido a propagação de erros no processo computacional. O teorema dos valores singulares oferece uma solução para esse problema.

O que faremos nessa seção não passa de uma interpretação em termos matriciais dos resultados obtidos na seção 7.5.

Seja A a matriz que representa a aplicação linear $T : \mathbb{R}^n \to \mathbb{R}^m$ com relação às bases canônicas do $\mathbb{R}^n$ e $\mathbb{R}^m$. Se $\mathcal{B} = \{v_1, \ldots, v_n\}$ é a base ortonormal do $\mathbb{R}^n$ formada por autovetores de A^*A, então $P_{\mathcal{E}}^{\mathcal{B}} = P$ é a matriz

$$P = (v_1 \ v_2 \ \ldots \ v_n)$$

cujas colunas são os vetores da base $\mathcal{B}$. Denotamos $Q = Q_{\mathcal{C}}^{\mathcal{E}}$ a matriz mudança da base $\mathcal{C}$ do $\mathbb{R}^m$ para a base canônica desse espaço e $D = T_{\mathcal{B}}^{\mathcal{C}}$. Então

$$D = QAP.$$

Como as matrizes P e Q são ortogonais, temos $A = Q^*DP^*$. Mudando a notação, temos

$$A = QDP,$$

chamada decomposição em **valores singulares** da matriz A.

Exemplo 8.5.1 Seja

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

Para obter a decomposição de A em valores singulares, obtemos a matriz A^TA;

$$A^TA = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix},$$

cujos autovalores são $\lambda_1 = 4$ e $\lambda_2 = 0$. Os valores singulares de A são, portanto, $\sigma_1 = \sqrt{4} = 2$ e $\sigma_2 = \sqrt{0} = 0$. A matriz P, cujas colunas são os autovetores normalizados de A^TA é

$$P = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

O vetor w_1 é dado por

$$w_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

Para obtermos os vetores w_2 e w_3, achamos uma base ortonormal de $\ker A^T$ (nesse exemplo, não é necessário utilizar o processo de ortogonalização de Gram-Schmidt):

$$w_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \quad \text{e} \quad w_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Portanto,

$$A = QDP = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix}.$$

◀

8.6 Exercícios

1. Demonstre a proposição 8.1.1 com relação às operações elementares (i) e (ii).

2. Mostre que toda matriz elementar tem inversa.

3. Mostre que o produto de matrizes triangulares inferiores (respectivamente, superiores) é uma matriz triangular inferior (resp., superior).

4. Dê um exemplo mostrando que é possível ter $A = LU = L'U'$, com L, L' matrizes triangulares inferiores com elementos diagonais todos iguais a 1 e U, U' matrizes triangulares superiores. (Compare com a proposição 8.1.3).

5. Seja A uma matriz simétrica invertível. Mostre que A^2 é uma matriz positiva-definida.

6. Suponhamos que o polinômio característico de $T : V \to V$ se decomponha em polinômios irredutíveis de grau um (quer dizer, todas as raízes do polinômio característico estão no corpo $\mathbb{K}$). Mostre[2]:

   (a) Se λ é autovalor de T, então $\overline{\lambda}$ é autovalor de $T^* : V \to V$;

   (b) Seja v um autovetor unitário associado ao autovalor $\overline{\lambda}$ de T^*. Decompondo $V = W \oplus W^{\perp}$, mostre que $T(W^{\perp}) \subset W^{\perp}$;

   (c) Supondo que o teorema de Schur seja válido em espaços de dimensão $n-1$ (cujas raízes do polinômio característico estão em $\mathbb{K}$), considere a restrição $T|_{W^{\perp}}$ - que produz uma base ortonormal $\mathcal{C}$ na qual a representação dessa restrição é triangular superior - e verifique que $T_{\mathcal{B}}$ é uma matriz triangular superior, sendo $\mathcal{B} = \mathcal{C} \cup \{v\}$.

7. Seja $\mathcal{B}$ uma base ortonormal de V. Suponhamos que a representação $A = T_{\mathcal{B}}$ do operador linear $T : V \to V$ seja uma matriz triangular superior. Mostre que T é normal se, e somente se, A é diagonal. Deduza o teorema 7.4.4 ao aplicar o teorema de Schur ao resultado anterior.

8. Na decomposição de Schur $U^*AU = T$ há unicidade da matriz triangular superior T?

9. A decomposição em valores singulares $A = QDP$ é única? Os valores singulares são únicos?

10. Quais são as diagonalizações ortogonais de A^*A e AA^*?

11. Seja A uma matriz $m \times n$. O posto de A é igual ao número de autovalores não nulos, contados de acordo com a multiplicidade?

[2]Compare com o teorema 6.5.4.

Referências Bibliográficas

[1] H. Anton and C. Rorres: Elementary Linear Algebra: Applications version, 6th. edition, Wiley, New York, 1991.

[2] R. Bellman: Introduction to Matrix Analysis, McGraw-Hill Book Company, New York, 1960. (Republicado na série “Classics in applied mathematics”, SIAM, 1995).

[3] H. P. Bueno: Equações Diferenciais Ordinárias - 1a. parte. Notas de aula de um curso. Departamento de Matemática da UFMG, 2001.

[4] N. Dunford and J. T. Schwarz: Linear operators I, Interscience, New York, 1968.

[5] Gantmacher, F. R.: The Theory of Matrices, vol. 1 and 2, Chelsea Publishing Co., New York, 1959.

[6] G. Golub e Charles Van Loan: Matrix Computations, 2nd. Edition, Johns Hopkins, Baltimore, 1989.

[7] M. Hirsch and S. Smale: Differential Equations, Dynamical Systems and Linear Algebra, Academic Press, New York, 1974.

[8] K. Hoffman e R. Kunze: Álgebra Linear, 2a. edição, Livros Técnicos e Científicos Editora, Rio de Janeiro, 1979.

[9] S. Lang: Linear Algebra, 3rd. Edition, Springer-Verlag, New York, 1989.

[10] P. D. Lax: Linear Algebra, Wiley-Interscience Publication, New York, 1997.

[11] S. J. Leon: Álgebra Linear com Aplicações, 4a. Edição, LTC Editora, Rio de Janeiro, 1999.

[12] E. L. Lima: Álgebra Linear, 2a. edição, IMPA, Rio de Janeiro, 1996.

[13] M. Montenegro: Notas de aula de um curso de Álgebra Linear, UFMG, 2000.

[14] M. Reed and B. Simon: Methods of Modern Mathematical Physics, vol. I, Academic Press, New York, 1972.

[15] R. J. Santos: Geometria Analítica e Álgebra Linear, Parte II, UFMG, 2000.

[16] J. Sotomayor: Lições de Equações Diferenciais Ordinárias, IMPA, Rio de Janeiro, 1979.

[17] G. Strang: Linear Algebra and its Applications, 3rd. edition, Harcourt, Fort Worth, 1988.

Índice Remissivo