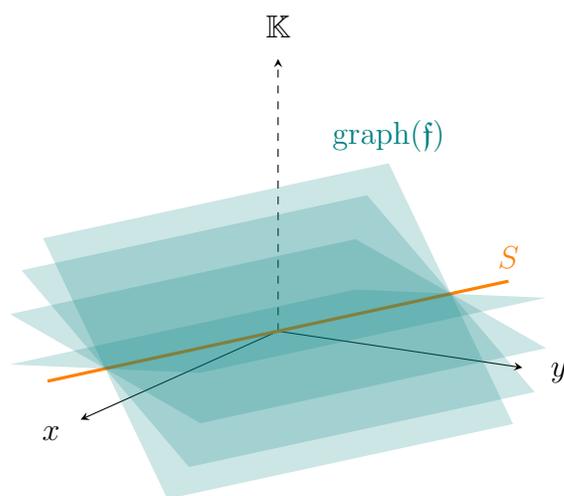


# Álgebra Linear Avançada

4 de janeiro de 2024

Ricardo M. S. Rosa

Instituto de Matemática  
Universidade Federal do Rio de Janeiro





## Sumário

Prefácio	11
Capítulo 1. Espaços Vetoriais	13
1. Espaços vetoriais	13
1.1. Definição e algumas propriedades	13
1.2. Exemplos de espaços vetoriais	17
1.3. Isomorfismo entre espaços vetoriais	23
1.4. Subespaço vetorial	26
1.5. Combinação linear	29
1.6. Subespaço gerado e conjunto gerador	31
2. Base e dimensão	33
2.1. Conjuntos linearmente dependentes e linearmente independentes	33
2.2. Subespaços linearmente independentes e soma direta	34
2.3. Base finita	36
2.4. Dimensão	38
2.5. Representação vetorial em uma base	44
2.6. Espaço complementar e co-dimensão	48
2.7. Espaço quociente a um subespaço	49
3. Base de Hamel	52
3.1. Combinação linear de elementos de um conjunto qualquer	52
3.2. Conjuntos linearmente dependentes e independentes	53
3.3. Base de Hamel	53
4. Aplicações	54
4.1. Determinantes e n-volumes	54
4.2. Representação computacional de matrizes especiais	58
5. Exercícios	63
Capítulo 2. Funcionais Lineares e Dualidade	71
1. Funcionais lineares e o dual algébrico	71
1.1. Definição	71
1.2. Exemplos de funcionais lineares	72

1.3. Produto de dualidade	74
2. Núcleo de um funcional linear e hiperplanos	75
2.1. O núcleo como um subespaço	75
2.2. Exemplos de núcleos	76
2.3. Co-dimensão do núcleo de um funcional não nulo	77
2.4. Espaço quociente ao núcleo	78
2.5. Caracterização de hiperplanos	79
2.6. Propriedades úteis	80
2.7. Caracterização de hiperplanos afim	81
3. Dimensão do espaço dual	82
3.1. Funcionais que identificam os vetores de uma base	83
3.2. Base dual e dimensão	83
3.3. Representação de vetores e funcionais lineares	85
3.4. Realização de um dual	89
3.5. Representação do espaço e do dual	91
4. O bidual	94
4.1. Definição	94
4.2. Morfismo canônico	94
4.3. Isomorfismo natural com o bidual em dimensão finita	95
5. Anulador de um subespaço	96
5.1. O subespaço anulador	96
5.2. Dimensão do anulador	97
6. Aplicações	98
6.1. Quadratura de polinômios	98
6.2. Subdiferenciais de funções convexas	100
6.3. Classificação de dados via máquina de vetores de suporte	100
7. Exercícios	100
Capítulo 3. Transformações Lineares	103
1. Fundamentos	103
1.1. Definição de transformação linear	103
1.2. Alguns conceitos essenciais	103
1.3. Exemplos	104
1.4. Transformação linear definida a partir da ação em uma base	109
2. O Teorema do Núcleo e da Imagem	110
2.1. Enunciado e demonstração	110
2.2. Consequências imediatas	112
3. O espaço das transformações lineares	113

3.1. O espaço vetorial das transformações lineares	113
3.2. Invertibilidade de transformações lineares	115
3.3. Representação matricial em espaços de dimensão finita	117
4. Transformação adjunta	121
4.1. Adjunta de uma transformação	122
4.2. Representação matricial de transformações adjuntas	123
4.3. Relações entre imagem e núcleo de uma transformação e de sua adjunta	125
5. Aplicações	127
5.1. Sistemas de equações lineares, forma escalonada e decomposição LU	127
5.2. Diferenciação automática	131
5.3. Análise dimensional e o Teorema de Buckingham-Pi	138
5.4. Dualidade em programação linear	144
5.5. Problemas primal e dual em máquina de vetores de suporte	145
6. Exercícios	145
Capítulo 4. Operadores Lineares	151
1. Fundamentos	151
1.1. Definição de operador linear	151
1.2. A álgebra dos operadores lineares	151
1.3. Operadores lineares invertíveis	154
1.4. Mudança de base	156
2. Similaridade	157
2.1. Operadores similares	158
2.2. Similaridade como relação de equivalência	159
2.3. Similaridade entre matrizes	161
3. Determinante de matrizes e operadores linear	163
3.1. Determinante de matrizes quadradas	163
3.2. Determinantes de matrizes especiais	165
3.3. Propriedades do determinante	166
3.4. O determinante de matrizes invertíveis	167
3.5. Calculando o determinante via escalonamento	168
3.6. Determinante invariante por similaridade	169
3.7. Determinante de operadores lineares	169
3.8. Determinante do operador adjunto	170
4. O operador projeção	171
4.1. Definição e exemplos	171
4.2. Propriedades do operador projeção	171
4.3. Caracterização de um operador projeção	173

4.4.	Representação matricial de um operador projeção	174
4.5.	Operadores involutivos	174
5.	Subespaços invariantes	176
5.1.	Definição	177
5.2.	Conexão com o operador projeção	178
5.3.	Representação matricial em blocos	180
5.4.	Subespaços encaixantes e forma triangular em blocos	183
5.5.	Subespaços invariantes independentes e forma diagonal em blocos	185
6.	Aplicações	186
6.1.	Transformada discreta de Fourier	186
6.2.	Operador de Koopman em sistemas dinâmicos	186
7.	Exercícios	187
Capítulo 5.	Operadores Nilpotentes	189
1.	Fundamentos	189
1.1.	Definição	189
1.2.	Exemplos	189
1.3.	Propriedades fundamentais	190
2.	Ciclos	191
2.1.	Cadeias de vetores e ciclos	191
2.2.	Exemplos de ciclos em operadores nilpotentes	192
2.3.	Caracterização de ciclos em operadores nilpotentes	197
3.	Forma normal de operados nilpotentes	198
3.1.	Forma normal no caso operadores nilpotentes cíclicos	198
3.2.	Forma normal no caso geral	199
3.3.	Mais exemplos	206
3.4.	Demonstrações alternativas	210
4.	Aplicações	212
4.1.	O sistema massa-mola com amortecimento crítico	212
5.	Exercícios	213
Capítulo 6.	Teoria Espectral e a Forma Normal de Jordan	217
1.	Fundamentos	217
1.1.	Autovalores e autovetores	217
1.2.	Autoespaços	218
1.3.	Espectro de um operador	220
2.	Diagonalização no caso de um conjunto completo de autovetores	221
2.1.	Autoespaços de autovalores distintos	221
2.2.	Diagonalização	223

3. Autovalores complexos e forma triangular superior	226
3.1. Existência de autovalores complexos	227
3.2. Forma triangular superior	230
3.3. Existência da forma triangular superior	231
3.4. Autovalores e a diagonal da forma triangular	233
4. A forma canônica de Jordan	236
4.1. Autoespaços generalizados	236
4.2. Polinômios anuladores	241
4.3. Multiplicidade dos autovalores e dimensão do autoespaço generalizado	245
4.4. Multiplicidade algébrica e multiplicidade geométrica dos autovalores	249
4.5. Forma canônica de Jordan	251
4.6. Exemplos	253
4.7. Forma canônica real de Jordan	256
5. Aplicações	260
5.1. Sistemas de equações diferenciais lineares homogêneas e a exponencial de matrizes	260
6. Exercícios	265
Capítulo 7. Norma e Produto Interno	269
1. Norma	269
1.1. Definição e exemplos	269
1.2. Topologia de um espaço vetorial normado	273
1.3. Norma de operadores e continuidade de transformações lineares	282
2. Produto interno real	291
2.1. Definição e exemplos	291
2.2. Norma proveniente do produto interno	293
3. Ortogonalidade em espaços vetoriais reais	296
3.1. Vetores e conjuntos ortogonais	296
3.2. Projeção ortogonal e o processo de Gram-Schmidt	298
3.3. Base ortonormal	304
4. Realização do dual e da adjunta via produto interno real	306
4.1. Realização do dual via produto interno	306
4.2. Realização da transformação adjunta via produto interno	310
5. Aplicações	312
5.1. Revisitando a exponencial de operadores	312
5.2. Número de condicionamento e análise de sensibilidade	312
5.3. Séries de Fourier e aproximações espectrais	315
5.4. Polinômios de Chebyshev e aproximações	320

5.5. O método de mínimos quadrados	321
6. Exercícios	322
Capítulo 8. Teoria Espectral em Espaços Vetoriais Reais com Produto Interno	325
1. Operadores especiais com simetria em relação à adjunta	325
1.1. Caracterização de projeção ortogonal como projeção autoadjunta	326
1.2. Isometria e operadores ortogonais	327
1.3. Operadores normais e relações de comprimento	329
2. Análise espectral de operadores autoadjuntos	331
2.1. Existência de autovalores reais	331
2.2. Invariância do ortogonal de um autoespaço	336
2.3. Diagonalização em base ortonormal	336
3. Decomposição espectral de operadores anti-autoadjuntos	338
3.1. Exemplos	338
3.2. Existência de subespaço invariante fundamental	339
3.3. Invariância do ortogonal de um subespaço invariante	341
3.4. Decomposição espectral	341
4. Decomposição espectral de operadores normais	343
4.1. Exemplos	343
4.2. Decomposição espectral	343
5. Decomposição em valores singulares (SVD)	346
5.1. Definição e exemplos	347
5.2. Demonstração de existência de decomposição SVD	351
5.3. Caracterização minimax	353
6. Aplicações	353
6.1. Cadeias de Markov	353
6.2. Algoritmo de PageRank	353
6.3. Decomposição ortogonal própria ou análise de componentes principais	353
6.4. Rigidez em equações diferenciais	354
7. Exercícios	356
Capítulo 9. Teoria Espectral em Espaços Vetoriais Complexos com Produto Interno	359
1. Produto interno em espaços vetoriais complexos	359
1.1. Definição e exemplos	359
1.2. Norma proveniente do produto interno complexo	360
1.3. Ortogonalidade	361
1.4. Realização do dual e da adjunta nos complexos	364
1.5. Operadores especiais	367

2. Decomposição de Schur e a triangularização ortogonal	367
2.1. Triangularização via subespaços invariantes ortogonais encaixantes	368
2.2. Decomposição	369
2.3. Propriedades e exemplos	370
3. Diagonalização de operadores normais	372
3.1. Operadores diagonalizáveis em alguma base ortonormal são normais	372
3.2. Todo operador normal é diagonalizável em alguma base ortonormal	374
3.3. Equivalência	377
4. Diagonalização de operadores hermitianos	378
5. Operadores anti-hermitianos	378
6. Aplicações	378
6.1. Fórmula de Cauchy, resíduos e outras ferramentas de análise complexa em teoria espectral	378
6.2. Séries de Fourier nos complexos	378
6.3. Modelos em mecânica quântica	378
7. Exercícios	378
Referências Bibliográficas	379
Símbolos e Abreviações	381
Índice Remissivo	383



## Prefácio

Mais um livro sobre Álgebra Linear Avançada!? Desnecessário, não? Pois é... Algum *insight* que não exista em outro lugar? Provavelmente não. Algum motivo especial? Não, nenhum motivo especial. Eu apenas gosto de escrever. Gosto de escrever as minhas próprias notas e de organizar os assuntos de maneira própria, com um princípio, meio e fim coerentes para mim. Apenas isso. Esse é o meu *modus operandi*.

Estas são, então, as minhas notas de aula para a disciplina de *Álgebra Linear do Mestrado em Matemática Aplicada do IM/UFRJ*, com espelho nas graduações em *Matemática Aplicada* e em *Engenharia Matemática*, sob o título de *Álgebra Linear Avançada*.

Há vários bons livros que serviram de inspiração. O ponto de partida foram as notas do Prof. César Niche (IM/UFRJ), que ministrou essa disciplina umas duas ou três vezes, e que por sua vez foram fortemente baseadas no sólido livro do Peter Lax [8], com algumas pitadas mais modernas de Peter Olver & Chehrzad Shakiban [11], rico em exemplos. Outras boas fontes são o livro do Elon Lages Lima [9], com inúmeros exercícios, o clássico livro do Serge Lang [7] e o livro do Sheldon Axler [1], com algumas visões interessantes. Para aplicações, Ivan Savov [12] também é bastante rico e Helene Shapiro [13] também tem algumas boas aplicações.

Qualquer um desses livros é um guia apropriado para uma disciplina de Álgebra Linear Avançada. Recomendo, em particular, o do Lax e do Elon, com eventuais incursões nos outros citados acima. E, em paralelo, os convido a seguirem as notas atuais, para o fio da meada.

Rio de Janeiro, 8 de setembro de 2023,

Ricardo M. S. Rosa



## CAPÍTULO 1

### Espaços Vetoriais

O foco principal de Álgebra Linear são as transformações lineares. São funções com uma estrutura particular de linearidade. Para que isso faça sentido, é preciso ter essa estrutura nos espaços em que essas transformações agem. Esses espaços, com essa estrutura mínima, de adição de vetores e de multiplicação por escalar, são os espaços vetoriais. Esses espaços são o tema desse capítulo.

#### 1. Espaços vetoriais

**1.1. Definição e algumas propriedades.** Espaços vetoriais são compostos de um conjunto de “vetores” e um conjunto de “escalares” com operações de adição entre vetores e multiplicação de um escalar por um vetor, satisfazendo condições apropriadas de compatibilidade.

O conjunto de escalares deve ser um *corpo*, que é uma estrutura algébrica como a dos números reais, com operações de adição e multiplicação entre eles e propriedades adequadas de comutatividade, associatividade, distributividade e, ainda, com a existência de elementos neutros para cada operação e a existência de elementos inversos, exceto que o elemento neutro aditivo não possui inversa multiplicativa.

Exemplos de corpos algébricos são os reais  $\mathbb{R}$ , os complexos  $\mathbb{C}$  e os racionais  $\mathbb{Q}$ . Um outro corpo famoso é o dos números  $p$ -ádicos  $\mathbb{Q}_p$ , localizados entre os racionais e os reais e definidos pelo somatório  $s = \sum_{i=k}^{\infty} a_i p^i$ , onde  $p$  é um número primo. Esses são todos corpos infinitos, i.e. com infinitos elementos. Corpos também podem ser finitos. Exemplos de corpos finitos são os espaços quociente  $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z} = \{0, 1, \dots, p-1\}$ , com  $p$  primo, munido das operações de adição e multiplicação em  $\mathbb{Z}$ , módulo  $p$ .

Podemos trabalhar com espaços vetoriais em um corpo qualquer, mas, para nós, aqui, vamos considerar apenas os números reais  $\mathbb{R}$  e os complexos  $\mathbb{C}$  (veja [Definição 1.2](#)). Estes têm a vantagem, em certos aspectos, de serem *arquimedianos* (i.e. possuem uma ordem com a propriedade de que quaisquer elementos  $x, y > 0$ , existe um inteiro  $n$  tal que  $nx > y$ , propriedade esta que os corpos finitos e os  $p$ -ádicos não têm); e *completos* (i.e. seqüências de Cauchy convergem, propriedade esta que os

racionais não possuem). Os complexos também tem a vantagem de serem *algebricamente fechados*, ou seja, qualquer polinômio não constante sobre os complexos possui raiz, propriedade que os reais não têm. Isso será devidamente explorado.

Vamos, finalmente, à definição formal de espaço vetorial.

DEFINIÇÃO 1.1 (Espaço vetorial). *Um **espaço vetorial** sobre um corpo  $\mathbb{K}$  é um conjunto  $X$  munido de duas operações, de **adição** e **multiplicação por escalar**, sendo*

- (i) a adição de  $X \times X$  em  $X$ , levando elementos  $\mathbf{u}, \mathbf{v} \in X$  em um elemento  $\mathbf{u} + \mathbf{v} \in X$ ,
- (ii) e a multiplicação por escalar de  $\mathbb{K} \times X$  em  $X$ , levando um elemento  $\lambda \in \mathbb{K}$  e um elemento  $\mathbf{u} \in X$  em um elemento  $\lambda \mathbf{u} \in X$ ,

satisfazendo as seguintes propriedades:

- (1) associatividade da adição,

$$(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w}), \quad \forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in X;$$

- (2) comutatividade da adição,

$$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}, \quad \forall \mathbf{u}, \mathbf{v} \in X;$$

- (3) existência do elemento neutro da adição

$$\exists \mathbf{0} \in X, \mathbf{u} + \mathbf{0} = \mathbf{u}, \quad \forall \mathbf{u} \in X;$$

- (4) existência do inverso aditivo,

$$\forall \mathbf{u} \in X, \exists (-\mathbf{u}) \in X, \mathbf{u} + (-\mathbf{u}) = \mathbf{0};$$

- (5) associatividade da multiplicação por escalar,

$$\lambda(\mu \mathbf{u}) = (\lambda\mu)\mathbf{u}, \quad \forall \lambda, \mu \in \mathbb{K}, \forall \mathbf{u} \in X;$$

- (6) elemento neutro multiplicativo  $1 \in \mathbb{K}$  do corpo é neutro para multiplicação por escalar,

$$1 \cdot \mathbf{u} = \mathbf{u}, \quad \forall \mathbf{u} \in X,$$

- (7) distributividade da multiplicação por escalar,

$$(\lambda + \mu)\mathbf{u} = \lambda \cdot \mathbf{u} + \mu \cdot \mathbf{u}, \quad \forall \lambda, \mu \in \mathbb{K}, \forall \mathbf{u} \in X;$$

- (8) distributividade da adição,

$$\lambda \cdot (\mathbf{u} + \mathbf{v}) = \lambda \cdot \mathbf{u} + \lambda \cdot \mathbf{v}, \quad \forall \lambda \in \mathbb{K}, \forall \mathbf{u}, \mathbf{v} \in X;$$

Os elementos de  $X$  são chamados de **vetores** e os elementos de  $\mathbb{K}$ , de **escalares**. O elemento neutro da adição,  $\mathbf{0} \in X$ , é chamado de **vetor nulo**.

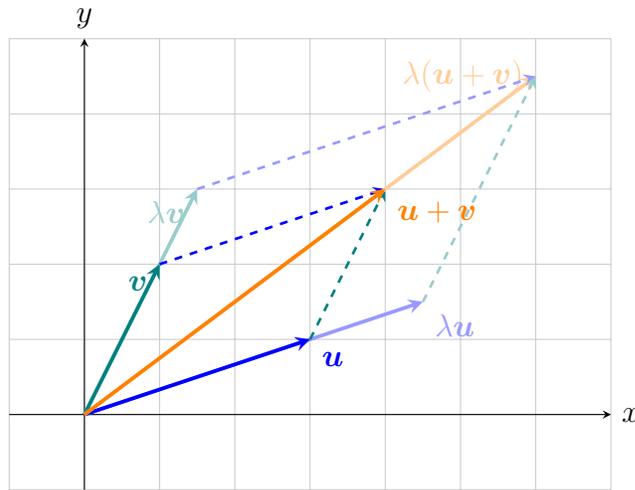


FIGURA 1.1. Ilustração da adição  $\mathbf{u} + \mathbf{v}$  de vetores, da multiplicação por escalar  $\lambda\mathbf{u}$ ,  $\lambda\mathbf{v}$  e da propriedade de distributividade  $\lambda(\mathbf{u} + \mathbf{v}) = \lambda\mathbf{u} + \lambda\mathbf{v}$ .

Como vamos nos restringir aos reais e aos complexos, fazemos a seguinte definição.

DEFINIÇÃO 1.2. *Pelas razões elencadas antes da Definição 1.1, vamos considerar, aqui, apenas espaços vetoriais sobre os reais ou sobre os complexos. Portanto, daqui para a frente,  $\mathbb{K} = \mathbb{R}$  ou  $\mathbb{C}$ . Em certos momentos, vamos especificar um ou outro, mas vários resultados valem para ambos e serão enunciados via  $\mathbb{K}$ . No caso específico  $\mathbb{K} = \mathbb{R}$ , chamaremos o espaço de **espaço vetorial real**, enquanto que no caso  $\mathbb{K} = \mathbb{C}$ , chamaremos o espaço de **espaço vetorial complexo**.*

OBSERVAÇÃO 1.1. Em termos de notação, vamos usar, com frequência, os símbolos  $X, Y, Z, \dots$  para denotar espaços vetoriais.

OBSERVAÇÃO 1.2. Alguns fatos:

- (1) *O vetor nulo é único.* De fato, se  $\mathbf{0}$  e  $\mathbf{0}'$  são neutros para a adição, então

$$\mathbf{0} = \mathbf{0} + \mathbf{0}' = \mathbf{0}'.$$

- (2) *O inverso aditivo é único.* De fato, dado  $\mathbf{u} \in X$ , se  $\mathbf{v}, \mathbf{w} \in X$  são tais que

$$\mathbf{u} + \mathbf{v} = \mathbf{0}, \quad \mathbf{u} + \mathbf{w} = \mathbf{0},$$

então

$$\mathbf{v} = \mathbf{v} + \mathbf{0} = \mathbf{v} + (\mathbf{u} + \mathbf{w}) = (\mathbf{v} + \mathbf{u}) + \mathbf{w} = \mathbf{0} + \mathbf{w} = \mathbf{w}.$$

- (3) *O elemento neutro do corpo neutraliza os vetores.* De fato, observe, primeiramente, que

$$\mathbf{u} = 1\mathbf{u} = (1 + 0)\mathbf{u} = 1\mathbf{u} + 0\mathbf{u} = \mathbf{u} + 0\mathbf{u},$$

de maneira que

$$\mathbf{0} = \mathbf{u} + (-\mathbf{u}) = (\mathbf{u} + 0\mathbf{u}) + (-\mathbf{u}) = (\mathbf{u} + (-\mathbf{u})) + 0\mathbf{u} = \mathbf{0} + 0\mathbf{u} = 0\mathbf{u},$$

ou seja,

$$0\mathbf{u} = \mathbf{0}, \quad \forall \mathbf{u} \in X.$$

- (4) *Inverso aditivo via multiplicação por escalar por  $-1$ .* Para todo  $\mathbf{u} \in X$ ,

$$\mathbf{0} = 0\mathbf{u} = (1 - 1)\mathbf{u} = 1\mathbf{u} + (-1)\mathbf{u} = \mathbf{u} + (-1)\mathbf{u}.$$

Logo, pela unicidade do inverso aditivo, segue que

$$(-1)\mathbf{u} = (-\mathbf{u}), \quad \forall \mathbf{u} \in X.$$

- (5) *Um espaço vetorial não pode ser vazio.* De fato, por definição, deve existir, pelo menos, o vetor nulo, i.e.  $\{\mathbf{0}\} \subset X$ .

OBSERVAÇÃO 1.3. Para simplificar a notação, escrevemos

$$\mathbf{u} + (-\mathbf{v}) = \mathbf{u} - \mathbf{v}, \quad \forall \mathbf{u}, \mathbf{v} \in X.$$

OBSERVAÇÃO 1.4. Espaços vetoriais são a base do Cálculo em Várias Variáveis. Por exemplo, para se tomar um derivada direcional  $\nabla_{\mathbf{v}}f(\mathbf{x})$  de uma função  $f : X \rightarrow Y$ , em um ponto  $\mathbf{x}$ , na direção de um vetor  $\mathbf{v}$ , calculamos o limite

$$\nabla_{\mathbf{v}}f(\mathbf{x}) = \lim_{h \rightarrow 0} \frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{h},$$

que envolve a multiplicação por escalar  $h\mathbf{v}$  e a adição  $\mathbf{x} + h\mathbf{v}$  no espaço  $X$  e a adição  $f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})$  e a multiplicação por escalar  $(1/h)(f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x}))$  no espaço  $Y$ . Ou seja, o domínio  $X$  e o contra-domínio  $Y$  da função  $f$  precisam ser espaços vetoriais (ou subconjuntos de espaços vetoriais ou, pelo menos, que tenham localmente uma estrutura parecida com a de espaços vetoriais, como nas variedades). A integração, como limite de somas de Riemann, também requer operações dessa forma. Ou seja, as estruturas vetoriais são fundamentais para dar sentido aos objetos de Cálculo.

**1.2. Exemplos de espaços vetoriais.** Vejamos, agora, alguns exemplos de espaços vetoriais.

EXEMPLO 1.1 (Espaço trivial). Por definição, um espaço vetorial não pode ser vazio, pois deve existir, pelo menos, o vetor nulo. Mas é permitido, pela definição, que o vetor nulo seja o único elemento do espaço vetorial. Um espaço vetorial  $X = \{\mathbf{0}\}$ , onde o vetor nulo é o único elemento do espaço é dito **espaço trivial**. Observe, ainda, que, dado um conjunto com um único elemento, digamos  $X = \{e\}$ , só há uma maneira possível de definir a adição e a multiplicação nesse espaço, a saber.  $e + e = e$  e  $\lambda e = e$ , de modo que  $e$  é o elemento neutro da adição vetorial e  $X = \{e\}$  é um espaço vetorial. Nesse caso,  $\mathbf{0} = e$  e podemos escrever  $X = \{\mathbf{0}\}$ .

EXEMPLO 1.2 (Corpo como espaço vetorial). Qualquer corpo pode ser visto como um espaço vetorial sobre si mesmo. Além disso, também pode ser visto como espaço vetorial sobre subcorpos dele mesmo. Por exemplo, podemos ver  $\mathbb{C}$  como um espaço vetorial sobre si mesmo ou como um espaço vetorial sobre os reais.

EXEMPLO 1.3 (Produto cartesiano finito). O produto cartesiano  $X = \mathbb{R}^d$  das  $d$ -uplas  $u = (x_1, \dots, x_d)$  munido da adição  $(x_1, \dots, x_d) + (y_1, \dots, y_d) = (x_1 + y_1, \dots, x_d + y_d)$  e da multiplicação por escalar  $\lambda(x_1, \dots, x_d) = (\lambda x_1, \dots, \lambda x_d)$  é um espaço vetorial sobre os reais  $\mathbb{R}$ . Por exemplo, no modelo Newtoniano, a posição de uma partícula é dada por suas coordenadas  $\mathbf{u} = (x, y, z) \in \mathbb{R}^3$ . O modelo epidemiológico SIR (Suscetíveis, Infectados e Recuperados) trata da evolução temporal das três componentes  $S = S(t)$ ,  $I = I(t)$  e  $R = R(t)$ , que pertencem ao espaço vetorial  $(S(t), I(t), R(t)) \in \mathbb{R}^3$ . Analogamente, podemos considerar  $\mathbb{C}^d$  como um espaço vetorial sobre  $\mathbb{C}$ , ou mesmo como um espaço vetorial sobre  $\mathbb{R}$ .

EXEMPLO 1.4 (Produto cartesiano com índice arbitrário). Mais geralmente, dado um corpo  $\mathbb{K}$  e um conjunto arbitrário  $A$ , podemos considerar o produto cartesiano  $\mathbb{K}^A$  formado pelo conjunto das funções

$$\mathbb{K}^A = \{x : A \rightarrow \mathbb{K}\},$$

cujos elementos podemos denotar como na forma de coordenadas

$$x = (x_a)_{a \in A},$$

onde

$$x_a = x(a).$$

O conjunto  $A$  é chamado de *conjunto dos índices*. O conjunto  $\mathbb{K}^A$  tem uma estrutura natural de espaço vetorial dada pelas operações “coordenada a coordenada”

$$(x_a)_a + (y_a)_a = (x_a + y_a)_a$$

e

$$\lambda(x_a)_a = (\lambda x_a)_a.$$

Por exemplo,  $\mathbb{R}^d$  pode ser visto como um conjunto dessa forma, com o conjunto de índices dado por  $A = \{1, \dots, d\}$ . O conjunto de todas as sequências  $(x_n)_{n \in \mathbb{N}}$  de números reais pode ser visto como um espaço vetorial  $\mathbb{R}^{\mathbb{N}}$ . O conjunto das funções  $f : I \subset \mathbb{R} \rightarrow \mathbb{R}$  pode ser visto como um espaço vetorial  $\mathbb{R}^I$ .

EXEMPLO 1.5 (Produto cartesiano de dois espaços vetoriais sobre o mesmo corpo). Sejam  $X$  e  $Y$  dois espaços vetoriais sobre o mesmo corpo  $\mathbb{K}$ . Então o produto cartesiano

$$X \times Y = \{(\mathbf{u}, \mathbf{v}); \mathbf{u} \in X, \mathbf{v} \in Y\}$$

munido das operações de adição e multiplicação por escalar definidas por

$$(\mathbf{u}_1, \mathbf{v}_1) + (\mathbf{u}_2, \mathbf{v}_2) = (\mathbf{u}_1 + \mathbf{u}_2, \mathbf{v}_1 + \mathbf{v}_2)$$

e

$$\lambda(\mathbf{u}, \mathbf{v}) = (\lambda\mathbf{u}, \lambda\mathbf{v})$$

é, também, um espaço vetorial sobre  $\mathbb{K}$ .

EXEMPLO 1.6 (Produto cartesiano de espaços vetoriais arbitrários sobre o mesmo corpo). Agora, se  $\{X_a\}_{a \in A}$  é uma família de espaços vetoriais sobre um mesmo corpo  $\mathbb{K}$ , então o produto cartesiano

$$X = \prod_{a \in A} X_a$$

é um espaço vetorial munido das operações de adição e multiplicação por escalar definidas por

$$(u_a)_a + (v_a)_a = (u_a + v_a)_a,$$

e

$$\lambda(u_a)_a = (\lambda u_a)_a,$$

onde, para cada índice  $a$ , a adição e a multiplicação por escalar ocorrem no espaço  $X_a$  correspondente.

EXEMPLO 1.7 (Conjunto versus espaço vetorial). Podemos distinguir o conjunto dos pontos em um plano  $xy$  do espaço vetorial associado. De outra forma, podemos considerar  $\mathbb{R}^2 = \{(x, y); x, y \in \mathbb{R}\}$  apenas como um conjunto de pontos ou então como um espaço vetorial quando munido das operações de soma de vetores e multiplicação por escalar. Dados dois pontos quaisquer  $P_1 = (x_1, y_1)$ ,  $P_2 = (x_2, y_2)$ , no plano, podemos considerar o segmento de reta

$$\overline{P_1 P_2} = \{(sx_1 + (1-s)x_2, sy_1 + (1-s)y_2); 0 \leq s \leq 1\},$$

assim como o vetor

$$\overrightarrow{P_1P_2} = (x_2 - x_1, y_2 - y_1).$$

Nesse sentido, podemos, também, distinguir a origem  $O = (0, 0)$ , no plano, do vetor nulo  $\mathbf{0} = \overrightarrow{OO}$ . Aqui, usamos  $\mathbb{R}^2$  para representar tanto o conjunto de pontos no plano  $xy$ , sem estrutura algébrica, como o espaço vetorial obtido a partir dele, juntando as operações de adição e multiplicação por escalar. Idem para  $\mathbb{R}^n$ ,  $n \in \mathbb{N}$ . Em alguns momentos fazemos distinção apenas entre os pontos e os vetores, mas apenas para efeitos didáticos. Mais para a frente, também vamos considerar normas em  $\mathbb{R}^n$  e produtos internos. No caso particular do produto escalar  $\mathbf{u} \cdot \mathbf{v} = \sum_j u_j v_j$ , obtido a partir das coordenadas de  $\mathbf{u} = (u_1, \dots, u_n)$  e de  $\mathbf{v} = (v_1, \dots, v_n)$ , e da norma euclidiana  $\|\mathbf{u}\| = \sqrt{\sum_j u_j^2}$  associada a ele, alguns textos usam uma notação diferente, como  $\mathbb{E}^n$ , chamando-o de *espaço Euclidiano*, reservando  $\mathbb{R}^n$  para o conjunto ou para espaços vetoriais sem estrutura de norma e produto interno específicos. Não vamos fazer essa distinção aqui, mas é importante saber que são objetos matemáticos diferentes (o conjunto, o espaço vetorial, o espaço vetorial com uma norma específica e o espaço vetorial com um produto interno específico).

EXEMPLO 1.8 (Complexificação de um espaço real). Dado um espaço vetorial  $X$  sobre o corpo dos reais, podemos definir a *complexificação* de  $X$  como sendo o espaço vetorial complexo dos pares ordenados  $\mathbf{u} = (\mathbf{u}_r, \mathbf{u}_i) \in X \times X$ , que denotamos por

$$\mathbf{u} = \mathbf{u}_r + i\mathbf{u}_i,$$

munidos das operações

$$\mathbf{u} + \mathbf{v} = (\mathbf{u}_r + \mathbf{v}_r) + i(\mathbf{u}_i + \mathbf{v}_i),$$

para  $\mathbf{u} = \mathbf{u}_r + i\mathbf{u}_i$  e  $\mathbf{v} = \mathbf{v}_r + i\mathbf{v}_i$ , e

$$\zeta \mathbf{u} = (\alpha + i\beta)(\mathbf{u}_r + i\mathbf{u}_i) = (\alpha\mathbf{u}_r - \beta\mathbf{u}_i) + i(\beta\mathbf{u}_r + \alpha\mathbf{u}_i).$$

Deixamos a cargo do leitor verificar que essas operações satisfazem os axiomas de espaço vetorial ([Exercício 5.6](#))

EXEMPLO 1.9 (Espaço  $\ell^1$ ). O espaço  $\ell^1 = \ell^1(\mathbb{R})$  é o espaço das sequências  $u = (u_1, u_2, \dots) \in \mathbb{R}^{\mathbb{N}}$  de números reais satisfazendo

$$\|u\|_1 = \sum_{n \in \mathbb{N}} |u_n| < \infty.$$

Observe que  $\|\cdot\|_1$  é homogêneo, pois

$$\|su\|_1 = |s|\|u\|_1,$$

Em particular, temos  $su \in \ell^1$ , para todo real  $s \in \mathbb{R}$ .

Agora, se  $u, v \in \ell^1$ , segue que

$$\|u + v\|_1 = \sum_{n \in \mathbb{N}} |u_n + v_n| \leq \sum_{n \in \mathbb{N}} (|u_n| + |v_n|) = \sum_{n \in \mathbb{N}} |u_n| + \sum_{n \in \mathbb{N}} |v_n| = \|u\|_1 + \|v\|_1.$$

Isso mostra, em particular, que se  $u, v \in \ell^1$ , então  $u + v \in \ell^1$ .

Dessa forma, vemos que  $\ell^p$  é um espaço vetorial.

EXEMPLO 1.10 (Espaço  $\ell^p$ ,  $p > 0$ ). Mais geralmente, o espaço  $\ell^p = \ell^p(\mathbb{R})$ , para  $p > 0$ , é o espaço das sequências  $u = (u_1, u_2, \dots) \in \mathbb{R}^{\mathbb{N}}$  de números reais satisfazendo

$$\|u\|_p = \left( \sum_{n \in \mathbb{N}} |u_n|^p \right)^{1/p} < \infty.$$

Observe que  $\|\cdot\|_p$  é homogêneo, pois

$$\|su\|_p = |s| \|u\|_p,$$

Em particular, temos  $su \in \ell^p$ , para todo real  $s \in \mathbb{R}$ .

Agora, se  $u, v \in \ell^p$ , segue, da desigualdade

$$(a + b)^p \leq (2 \max\{a, b\})^p = 2^p \max\{a^p, b^p\} \leq 2^p (a^p + b^p), \quad \forall a, b \geq 0, \forall p > 0,$$

que

$$\|u + v\|_p^p \leq 2^p (\|u\|_p^p + \|v\|_p^p),$$

portanto

$$\|u + v\|_p \leq 2 (\|u\|_p^p + \|v\|_p^p)^{1/p}.$$

Isso mostra, em particular, que se  $u, v \in \ell^p$ , então  $u + v \in \ell^p$ . Dessa forma, vemos que  $\ell^p$  é um espaço vetorial.

EXEMPLO 1.11 (Espaço  $\ell^\infty$ ). Para completar, temos o espaço  $\ell^\infty = \ell^\infty(\mathbb{R})$ , que é o espaço das sequências  $u = (u_1, u_2, \dots) \in \mathbb{R}^{\mathbb{N}}$  de números reais satisfazendo

$$\|u\|_\infty = \max_{n \in \mathbb{N}} |u_n| < \infty.$$

Deixamos para o leitor os detalhes da verificação de que isso forma um espaço vetorial.

EXEMPLO 1.12 (Polinômios). O espaço

$$\mathcal{P}_n(\mathbb{K}) = \{a_0 + a_1x + \dots + a_nx^n; a_0, \dots, a_n \in \mathbb{K}\}$$

de polinômios de grau no máximo  $n \in \mathbb{N}$  em uma variável  $x$  sobre um corpo  $\mathbb{K}$  é um espaço vetorial munido das operações de adição e multiplicação por escalar dadas respectivamente por

$$\begin{aligned}(p + q)(x) &= p(x) + q(x) = (a_0 + a_1x + \cdots + a_nx^n) + (b_0 + b_1x + \cdots + b_nx^n) \\ &= (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_n + b_n)x^n\end{aligned}$$

e

$$(\lambda p)(x) = \lambda p(x) = \lambda(a_0 + a_1x + \cdots + a_nx^n) = \lambda a_0 + \lambda a_1x + \cdots + \lambda a_nx^n$$

Analogamente, o espaço

$$\mathcal{P}(\mathbb{K}) = \{a_0 + a_1x + \cdots + a_nx^n; a_0, \dots, a_n \in \mathbb{K}, n \in \mathbb{N}\}$$

de polinômios de grau qualquer também é um espaço vetorial quando munido da mesma operação de multiplicação por escalar e com a adição feita de maneira análoga, mesmo com polinômios de diferentes graus, apenas somando os monômios de mesmo grau e acrescentando os monômios que estão presentes em apenas um dos polinômios, que é equivalente a completar com coeficientes nulos os monômios que estão em um polinômio e não estão em outros.

EXEMPLO 1.13 (Matrizes). O espaço  $\mathbb{K}^{m \times n}$  das matrizes  $m \times n$ , i.e.  $m$  linhas e  $n$  colunas, onde  $m, n \in \mathbb{N}$ , também é um espaço vetorial munido das operações realizadas “coeficiente a coeficiente”, i.e. se  $\mathbf{A} = (a_{ij})_{i=1, \dots, m, j=1, \dots, n}$ ,  $\mathbf{B} = (b_{ij})_{i=1, \dots, m, j=1, \dots, n}$  e  $\lambda \in \mathbb{K}$ , então

$$\mathbf{A} + \mathbf{B} = (a_{ij} + b_{ij})_{i=1, \dots, m, j=1, \dots, n}, \quad \lambda \mathbf{A} = (\lambda a_{ij})_{i=1, \dots, m, j=1, \dots, n}.$$

Uma matriz  $m \times n$  pode ser representada da seguinte forma

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

O elemento neutro da adição matricial é a matriz nula  $\mathbf{O} = (0)_{i,j=1}^{i=m, j=n}$ , em que todos os seus coeficientes são nulos. Para explicitar as dimensões, podemos escrever  $\mathbf{O} = \mathbf{O}_{m,n}$ , para indicar a matriz nula  $m \times n$ .

As matrizes ainda têm uma outra operação que é a de *multiplicação matricial*. Dadas duas matrizes  $\mathbf{A} \in \mathbb{K}^{m \times p}$  e  $\mathbf{B} \in \mathbb{K}^{p \times n}$ , com o número de colunas de  $\mathbf{A}$  coincidindo com o número de linhas de  $\mathbf{B}$ , podemos formar uma nova matriz  $\mathbf{AB} \in \mathbb{K}^{m \times n}$  multiplicando apropriadamente os seus coeficientes. De fato, se  $\mathbf{A} = (a_{ij})_{i=1, \dots, m, j=1, \dots, p}$  e

$\mathbf{B} = (b_{ij})_{i=1,\dots,p,j=1,\dots,n}$ , então  $\mathbf{AB} = (c_{ij})_{i=1,\dots,m,j=1,\dots,n}$  é definida por

$$c_{ij} = \sum_{k=1,\dots,n} a_{ik}b_{kj}.$$

Cada linha  $i$  de  $\mathbf{A}$  é multiplicada, coeficiente a coeficiente, por cada coluna  $j$  de  $\mathbf{B}$ , e os produtos são somados entre si, para gerar o elemento  $c_{ij}$  de  $\mathbf{AB}$ . Analogamente, podemos multiplicar uma matriz  $\mathbf{A} = (a_{ij})_{i=1,\dots,m,j=1,\dots,p}$  por um vetor  $\mathbf{u} = (x_j)_{j=1,\dots,p}$  nos dando o vetor  $\mathbf{Au} = (y_i)_{i=1,\dots,m}$ , com

$$y_i = \sum_{j=1,\dots,p} a_{ij}x_j.$$

A multiplicação matriz-vetor  $\mathbf{Au}$  está ligada a transformações lineares, enquanto que o produto matriz-matriz está ligado à composição de transformações lineares, conforme veremos no [Capítulo 3](#). Uma matriz com propriedades especiais para a multiplicação é a matriz identidade  $\mathbf{I} = \mathbf{I}_n = (\delta_{ij})_{i,j=1}^n$ , onde  $\delta_{ij}$  é o *Delta de Kronecker*, definido por  $\delta_{ij} = 1$ , se  $i = j$ , e  $\delta_{ij} = 0$ , se  $i \neq j$ . Em representação matricial,

$$\mathbf{I} = \mathbf{I}_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix}$$

EXEMPLO 1.14. Variedades diferenciáveis  $\mathcal{M}$  são estruturas que podem ser pensadas localmente como um espaço vetorial [3]. A variedade em si não é um espaço vetorial, mas, em cada ponto  $p \in \mathcal{M}$ , podemos definir um *espaço tangente*  $T_p\mathcal{M}$ . Este, sim, é um espaço vetorial. O conjunto  $T\mathcal{M}$  de todos os espaços tangentes é o chamado *fibrado tangente* de  $\mathcal{M}$ .

EXEMPLO 1.15 (Registro de áudio). Em processamento de sinais digitais de áudio, cada canal é registrado como um (longo) vetor, cujo tamanho depende do tempo de gravação e da frequência de amostragem (*sampling*) (veja [Figura 1.2](#)). Uma frequência comum é a de 44,100 Hz, ou seja, 44,100 registros a cada segundo, contendo uma medida do nível de pressão acústica a cada registro. A variação desse nível ao longo do tempo é refletida em diversos elementos musicais, como altura, dinâmica e timbre. Essa frequência de amostragem é capaz de resolver todas as frequências audíveis pelos seres humanos, que vão de algo da ordem de 20 Hz a 20,000 Hz, evitando o efeito de *aliasing*, em que uma frequência mais alta é confundida com uma mais baixa, por falta de amostragem adequada (procure pelo Teorema de Nyquist-Shannon, que garante que cada frequência de amostragem de  $f$  Hz é capaz de registrar todas as frequências

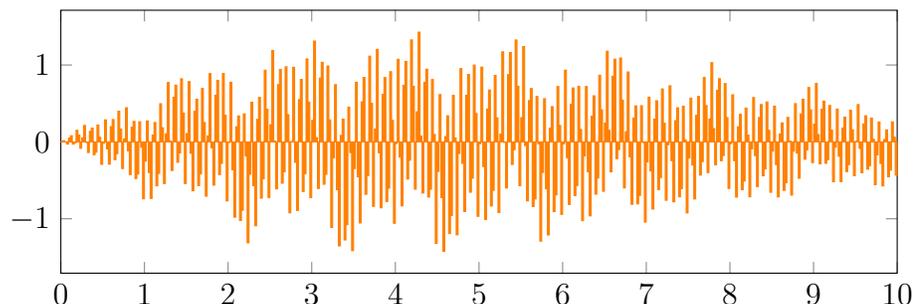


FIGURA 1.2. Trecho de 10 ms de um canal de áudio com frequência de amostragem 44,100 Hz, contendo 410 amostras, visualizado como um gráfico do sinal em função do tempo.

abaixo de  $(f/2)$  Hz). Teoricamente, podemos considerar esse sinal como um vetor em  $\mathbb{R}^n$ , para  $n = 44,100$  para cada segundo gravado. Porém, o computador não é capaz de armazenar um conjunto ilimitado de números reais. Cada amostra é armazenada, então, com um certo número de bits, tipicamente 16 ou 24 bits. Em cima disso, usa-se, ainda, técnicas de compressão de dados, para reduzir o tamanho do arquivo de áudio. De qualquer maneira, a idealização do registro de áudio como um vetor em um espaço  $\mathbb{R}^n$  é extremamente útil na análise de sinais de áudio. O processamento de áudio, como a análise e ajuste do espectro de frequências, pode ser visto como transformações nesses espaços, como a transformada discreta de Fourier, que veremos posteriormente. Além de vários outros efeitos em instrumentos musicais, como *delay*, *reverb* e distorção, assim como os efeitos de resposta de impulso (*impulse response*), para a simulação de timbres de outros instrumentos, amplificadores e ambientes.

EXEMPLO 1.16 (Imagens e vídeos digitais). Uma imagem digital também pode ser pensada como um elemento  $\mathbf{u} = (u_{ijk})_{ijk}$  de um espaço vetorial  $\mathbb{R}^{m \times n} \times \mathbb{R}^3$ , com  $k = 1, 2, 3$  representando as intensidades de cada cor, digamos em RGB (vermelho, verde e azul),  $j = 1, \dots, n$  o número de colunas e  $i = 1, \dots, m$  o número de linhas de cada imagem. Vídeos podem ser pensados como elementos em  $\mathbb{R}^{m \times n} \times \mathbb{R}^3 \times \mathbb{R}^k$ , com  $k$  representando o número de quadros registrados ao longo do tempo, que depende do intervalo de tempo e do número de quadros por segundo. O processamento de imagens e vídeos pode ser pensado como um conjunto de transformações (lineares ou não-lineares) nesses espaços.

**1.3. Isomorfismo entre espaços vetoriais.** Um polinômio de grau 2, com coeficientes reais, pode ser escrito, de maneira geral, na forma

$$p(x) = ax^2 + bx + c,$$

onde  $a, b, c \in \mathbb{R}$  são arbitrários. Tudo o que determina esse polinômio são os coeficientes  $a, b, c$ , que formam uma tripla  $(a, b, c)$  arbitrária em  $\mathbb{R}^3$ . De certa forma, um polinômio pode ser representado por um vetor em  $\mathbb{R}^3$ . Inclusive as operações de adição e multiplicação por escalar se correspondem, visto que

$$p_1(x) + p_2(x) = a_1x^2 + b_1x + c_1 + a_2x^2 + b_2x + c_2 = (a_1 + a_2)x^2 + (b_1 + b_2)x + (c_1 + c_2)$$

corresponde ao vetor  $(a_1 + a_2, b_1 + b_2, c_1 + c_2) = (a_1, b_1, c_1) + (a_2, b_2, c_2)$  e

$$\lambda p(x) = \lambda(ax^2 + bx + c) = (\lambda a)x^2 + (\lambda b)x + (\lambda c)$$

corresponde ao vetor  $(\lambda a, \lambda b, \lambda c) = \lambda(a, b, c)$ . Ou seja, do ponto de vista do conjunto e da estrutura linear do espaço, os espaços  $\mathcal{P}_2$  e  $\mathbb{R}^3$  são similares. Tudo o que se pode fazer com adições de elementos e multiplicações por escalar pode ser feito em um ou em outro, antes ou depois de fazer essa correspondência. Isso pode ser formalizado com o conceito de isomorfismo.

**DEFINIÇÃO 1.3.** *Um isomorfismo entre espaços vetoriais  $X$  e  $Y$  sobre o mesmo corpo  $\mathbb{K}$  é uma bijeção  $\varphi : X \rightarrow Y$  que preserva a estrutura linear dos espaços, i.e.*

$$\varphi(\mathbf{u} + \mathbf{v}) = \varphi(\mathbf{u}) + \varphi(\mathbf{v})$$

e

$$\varphi(\lambda \mathbf{u}) = \lambda \varphi(\mathbf{u}),$$

para  $\mathbf{u}, \mathbf{v} \in X$ ,  $\lambda \in \mathbb{K}$  quaisquer. Nesse caso, dizemos que  $X$  e  $Y$  são **isomorfos** entre si.

**OBSERVAÇÃO 1.5.** A definição acima de isomorfismo garante automaticamente que a inversa  $\varphi^{-1} : Y \rightarrow X$  também preserva a estrutura linear, de  $Y$  para  $X$ . Vale ressaltar, também que as adições e multiplicações por escalar em cada lado das equações acima são nos seus respectivos espaços.

**OBSERVAÇÃO 1.6.** “Isomorfismo” significa “mesma forma”, das raízes gregas “iso” (igual) e “morphus” (forma). É um conceito pervasivo em matemática, visando classificar estruturas que têm a mesma forma, e logicamente depende da estrutura do objeto em estudo. Um isomorfismo entre conjuntos basta ser uma *bijeção*, pois apenas os objetos contam. Um isomorfismo entre espaços topológicos deve ser uma *bijeção contínua*, que é a essência do conceito de topologia. Um isomorfismo entre grupos deve preservar a operação de multiplicação (ou adição ou seja lá qual for a operação binária do grupo). É assim por diante. No caso de espaços vetoriais, é a linearidade que deve ser preservada.

**EXEMPLO 1.17.** No caso dos polinômios de grau até dois, a função  $ax^2 + bx + c \mapsto (a, b, c)$  é um isomorfismo entre  $\mathcal{P}_2$  e  $\mathbb{R}^3$ .

EXEMPLO 1.18. No caso de uma matriz

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{R}^{2 \times 2},$$

um isomorfismo com  $\mathbb{R}^4$  é obtido levando a matriz com esses coeficientes no vetor  $(a, b, c, d) \in \mathbb{R}^4$ .

EXEMPLO 1.19. Caso tenhamos dois espaços da forma  $X = \mathbb{R}^n$  e  $Y = \mathbb{R}^m$ , então  $\mathbb{R}^n \times \mathbb{R}^m$  não é exatamente o mesmo que  $\mathbb{R}^{n+m}$ , mas os espaços são trivialmente isomorfos e são, muitas vezes, tratados indistintamente. No primeiro caso, temos  $\mathbf{v} \in X \times Y$  da forma  $\mathbf{v} = (\mathbf{u}, \mathbf{w})$  onde  $\mathbf{u} = (x_1, \dots, x_n)$  e  $\mathbf{w} = (y_1, \dots, y_m)$ , de maneira que

$$\mathbf{v} = ((x_1, \dots, x_n), (y_1, \dots, y_m)),$$

enquanto que  $\tilde{\mathbf{v}} \in \mathbb{R}^{n+m}$  tem a forma

$$\tilde{\mathbf{v}} = (z_1, \dots, z_n, z_{n+1}, \dots, z_{n+m}),$$

mas é claro que podemos fazer o isomorfismo

$$((x_1, \dots, x_n), (y_1, \dots, y_m)) \mapsto (x_1, \dots, x_n, y_1, \dots, y_m).$$

Exemplos naturais em que se considera  $\mathbb{R}^n \times \mathbb{R}^m$  ao invés de  $\mathbb{R}^{n+m}$  aparecem em mecânica e em muitos outros problemas. Por exemplo, o estado cinemático de uma única partícula se movendo no espaço é dado pelo par ordenado  $(\mathbf{x}, \mathbf{v}) \in \mathbb{R}^3 \times \mathbb{R}^3$ , onde  $\mathbf{x} = (x_1, x_2, x_3)$  indica a posição da partícula e  $\mathbf{v} = (v_1, v_2, v_3)$  indica a velocidade da partícula. O movimento de  $n$  partículas se escreve em um espaço  $\mathbb{R}^{3n} \times \mathbb{R}^{3n}$ , com  $n$  vetores posição e  $n$  vetores velocidade. Em mecânica Lagrangiana, usamos  $\mathbf{q} \in \mathbb{R}^k$  para as coordenadas generalizadas de um sistema de partículas, incluindo possivelmente alguma restrição, e  $\mathbf{p} \in \mathbb{R}^k$  para os momentos generalizados, formando o par  $(\mathbf{p}, \mathbf{q}) \in \mathbb{R}^k \times \mathbb{R}^k$ . Em estudos de sensibilidade nas condições iniciais  $\mathbf{x}_0$  de um sistema  $t \mapsto \mathbf{x}(t; \mathbf{x}_0)$  com  $\mathbf{x}(t; \mathbf{x}_0) \in \mathbb{R}^n$ , é comum considerarmos o sistema estendido envolvendo o estado  $\mathbf{x}(t; \mathbf{x}_0) \in U = \mathbb{R}^n$  do sistema no instante  $t$  junto com a evolução da diferencial  $D_{\mathbf{x}_0} \mathbf{x}(t; \mathbf{x}_0) \in \mathbb{R}^{n \times n}$ , nos dando um sistema em  $\mathbb{R}^n \times \mathbb{R}^{n \times n}$ , que é isomorfo a  $\mathbb{R}^{n+n^2}$ . Caso o sistema dependa de parâmetros  $\mathbf{p} \in \mathbb{R}^k$  e queiramos estudar a sensibilidade de  $x(t; \mathbf{x}_0, \mathbf{p})$  também nos parâmetros, devemos incluir as diferenciais  $(D_{\mathbf{x}_0} x(t; \mathbf{x}_0, \mathbf{p}), D_{\mathbf{p}} x(t; \mathbf{x}_0, \mathbf{p})) \in \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times k}$ , nos dando um sistema em  $\mathbb{R}^n \times \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times k}$ .

EXEMPLO 1.20 (Matrizes em blocos). De maneira análoga à feita no [Exemplo 1.19](#), podemos separar uma matriz  $\mathbb{R}^{m \times n}$  em blocos  $\mathbb{R}^{(m_1+m_2) \times n} = \mathbb{R}^{m_1 \times n} \times \mathbb{R}^{m_2 \times n}$  ou  $\mathbb{R}^{m \times (n_1+n_2)} = \mathbb{R}^{m \times n_1} \times \mathbb{R}^{m \times n_2}$  ou, mais geralmente,  $\mathbb{R}^{(m_1+m_2) \times (n_1+n_2)} = (\mathbb{R}^{m_1 \times n_1} \times$

$\mathbb{R}^{m \times n_2}) \times (\mathbb{R}^{m_2 \times n_1} \times \mathbb{R}^{m_2 \times n_2})$ . No primeiro caso, acompanhando a estrutura de linhas, escrevemos

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{m_1 \times n} \\ \mathbf{A}_{m_2 \times n} \end{bmatrix},$$

onde as dimensões dos blocos de matrizes estão indicadas no subscrito, i.e.  $\mathbf{A}_{m_1 \times n} \in \mathbb{R}^{m_1 \times n}$  e  $\mathbf{A}_{m_2 \times n} \in \mathbb{R}^{m_2 \times n}$ .

No segundo caso, escrevemos

$$\mathbf{A} = [\mathbf{A}_{m \times n_1} \quad \mathbf{A}_{m \times n_2}].$$

No terceiro caso, escrevemos

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{m_1 \times n_1} & \mathbf{A}_{m_1 \times n_2} \\ \mathbf{A}_{m_2 \times n_1} & \mathbf{A}_{m_2 \times n_2} \end{bmatrix}.$$

Quando as dimensões são compatíveis, podemos fazer os produtos matriciais bloco a bloco:

$$[\mathbf{A} \quad \mathbf{B}] \begin{bmatrix} \mathbf{C} \\ \mathbf{D} \end{bmatrix} = \mathbf{AC} + \mathbf{BD}, \quad \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} [\mathbf{C} \quad \mathbf{D}] = \begin{bmatrix} \mathbf{AC} & \mathbf{AD} \\ \mathbf{BC} & \mathbf{BD} \end{bmatrix},$$

além de

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{E} \\ \mathbf{F} \end{bmatrix} = \begin{bmatrix} \mathbf{AE} + \mathbf{BF} \\ \mathbf{CE} + \mathbf{DF} \end{bmatrix}, \quad [\mathbf{A} \quad \mathbf{B}] \begin{bmatrix} \mathbf{C} & \mathbf{D} \\ \mathbf{E} & \mathbf{F} \end{bmatrix} = [\mathbf{AC} + \mathbf{BE} \quad \mathbf{AD} + \mathbf{BF}],$$

e

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{bmatrix} = \begin{bmatrix} \mathbf{AE} + \mathbf{BG} & \mathbf{AF} + \mathbf{BH} \\ \mathbf{CE} + \mathbf{DG} & \mathbf{CF} + \mathbf{DH} \end{bmatrix}$$

É claro que essa mesma ideia pode ser generalizada para vários blocos:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1k} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{l1} & \mathbf{A}_{l2} & \cdots & \mathbf{A}_{lk} \end{bmatrix},$$

onde  $\mathbf{A}_{ij} \in \mathbb{R}^{m_i \times n_j}$  e  $\mathbf{A} \in \mathbb{R}^{m \times n} = \mathbb{R}^{(m_1 + \cdots + m_l) \times (n_1 + \cdots + n_k)}$ ,  $n = n_1 + \cdots + n_k$ ,  $m = m_1 + \cdots + m_l$ .

**1.4. Subespaço vetorial.** Subconjuntos de um espaço vetorial que herdam, de certa forma, a estrutura linear têm um papel importante na teoria e ganham um nome especial.

**DEFINIÇÃO 1.4 (Subespaço).** Um *subespaço* de um espaço vetorial  $X$  sobre um corpo  $\mathbb{K}$  é um subconjunto não-vazio  $S \subset X$  que é fechado por combinações lineares de seus vetores, ou seja, é tal que

- (1)  $\mathbf{u} + \mathbf{v} \in S$ , para todo  $\mathbf{u}, \mathbf{v} \in S$ ;  
 (2)  $\lambda \mathbf{u} \in S$ , para todo  $\lambda \in \mathbb{K}$  e todo  $\mathbf{u} \in S$ ;

Caso  $S$  seja diferente de  $X$ , ou seja, caso  $S$  esteja estritamente contido em  $X$ , dizemos que  $S$  é um **subespaço próprio** de  $X$  e escrevemos, mais precisamente,  $S \subsetneq V$ .

OBSERVAÇÃO 1.7. Graças às propriedades de invariância, um subespaço herda do espaço todo as operações de adição e multiplicação por escalar de tal forma que o próprio subespaço se torna um espaço vetorial. Observe que escolhendo um vetor  $v \in S$  qualquer (estamos assumindo que um subespaço é não-vazio) e tomando  $\lambda = 0$  em  $\mathbb{K}$ , temos que o vetor nulo  $0\mathbf{v} = \mathbf{0}$  pertence a  $S$ . Portanto, temos, naturalmente, que  $\{\mathbf{0}\} \subseteq S \subseteq V$ .

OBSERVAÇÃO 1.8. Em termos de notação, vamos usar, com frequência, os símbolos  $U, V, W, S, \dots$  para denotar subespaços vetoriais.

EXEMPLO 1.21 (Planos no espaço). Um plano  $P = \{(x, y, z) \in \mathbb{R}^3; ax + by + cz = d\}$ , onde  $a, b, c, d \in \mathbb{R}$ , é um subespaço vetorial de  $\mathbb{R}^3$  se, e somente se,  $d = 0$ , ou seja, se, e somente se,  $P$  passa pela origem. De fato, se  $\mathbf{u}_1 = (x_1, y_1, z_1)$  e  $\mathbf{u}_2 = (x_2, y_2, z_2)$  são elementos de  $P$ , então

$$\mathbf{u}_1 + \mathbf{u}_2 = (x_1 + x_2, y_1 + y_2, z_1 + z_2),$$

ou seja,  $\mathbf{u}_1 + \mathbf{u}_2 = (x, y, z)$  com  $x = x_1 + x_2$ ,  $y = y_1 + y_2$  e  $z = z_1 + z_2$ . Dessa forma, com

$$\begin{aligned} ax + by + cz &= a(x_1 + x_2) + b(y_1 + y_2) + c(z_1 + z_2) \\ &= ax_1 + by_1 + cz_1 + ax_2 + by_2 + cz_2 = d + d = 2d. \end{aligned}$$

Portanto,  $\mathbf{u}_1 + \mathbf{u}_2 \in P$  se, e somente se,  $2d = d$ , ou seja, se, e somente se  $d = 0$ . Por outro lado, se  $\mathbf{u} = (x, y, z) \in P$  e  $\lambda \in \mathbb{R}$  arbitrário, então  $\lambda \mathbf{u} = (\lambda x, \lambda y, \lambda z)$ . Temos,

$$a(\lambda x) + b(\lambda y) + c(\lambda z) = \lambda(ax + by + cz) = \lambda d,$$

de modo que  $\lambda \mathbf{u} \in P$  se, e somente se,  $\lambda d = d$ , portanto se, e somente se,  $d = 0$ . Com isso, concluímos que  $P$  é subespaço se, e somente se,  $d = 0$ .

EXEMPLO 1.22 (Polinômios). O conjunto  $\mathcal{P}_n(\mathbb{R})$  de polinômios de grau até  $n$  é subespaço de  $\mathcal{P}_m(\mathbb{R})$ , para  $m \geq n$ , além de ser subespaço do espaço  $\mathcal{P}(\mathbb{R})$  dos polinômios de grau arbitrário, assim como do espaço  $\mathcal{C}(\mathbb{R})$  das funções contínuas em  $\mathbb{R}$ , entre outros.

OBSERVAÇÃO 1.9 (Matriz transposta). Para os próximos dois exemplos, lembramos que, dada uma matriz quadrada  $\mathbf{A} = (a_{ij})_{i,j=1}^n$  em  $\mathbb{R}^{n \times n}$ ,  $n \in \mathbb{N}$ , a *transposta*  $\mathbf{A}^{\text{tr}}$

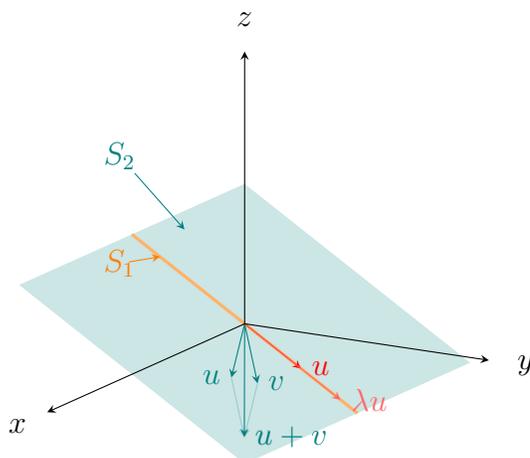


FIGURA 1.3. Exemplos de subespaços:  $S_1$ , de dimensão um, e  $S_2$ , de dimensão 2, com as ilustrações de serem invariantes por multiplicação por escalar e adição.

de  $\mathbf{A}$  é a matriz quadrada  $\mathbf{A}^{\text{tr}} = (\tilde{a}_{ij})_{i,j=1}^n$  obtida trocando-se as linhas pelas colunas, i.e.  $\tilde{a}_{ij} = a_{ji}$ ,  $i, j = 1, \dots, n$ .

EXEMPLO 1.23 (Matrizes simétricas). O conjunto

$$S = \{\mathbf{A} \in \mathbb{R}^{n \times n}; \mathbf{A}^{\text{tr}} = \mathbf{A}\}$$

das matrizes simétricas é um subespaço de  $\mathbb{R}^{n \times n}$ .

EXEMPLO 1.24 (Matrizes com traço nulo). Dada uma matriz quadrada  $\mathbf{A} = (a_{ij})_{i,j=1}^n$  em  $\mathbb{R}^{n \times n}$ ,  $n \in \mathbb{N}$ , a *traço*  $\text{tr}(\mathbf{A})$  de  $\mathbf{A}$  é o somatório dos elementos da diagonal de  $\mathbf{A}$ , i.e.

$$\text{tr}(\mathbf{A}) = a_{11} + \dots + a_{nn}.$$

O conjunto

$$S = \{\mathbf{A} \in \mathbb{R}^{n \times n}; \text{tr}(\mathbf{A}) = 0\}$$

das matrizes de traço nulo é um subespaço de  $\mathbb{R}^{n \times n}$ .

EXEMPLO 1.25 (Matrizes anti-simétricas). O conjunto

$$A = \{\mathbf{A} \in \mathbb{R}^{n \times n}; \mathbf{A}^{\text{tr}} = -\mathbf{A}\}$$

das matrizes anti-simétricas também é um subespaço de  $\mathbb{R}^{n \times n}$ . Em termos dos componentes, sendo  $\mathbf{A} = (a_{ij})_{i,j=1}^n$  em  $\mathbb{R}^{n \times n}$  anti-simétrica, com  $n \in \mathbb{N}$ , temos  $a_{ji} = -a_{ij}$ ,

para todo  $1 \leq i, j \leq n$ . Em particular, temos  $a_{ii} = -a_{ii}$ , ou seja,  $a_{ii} = 0$ , para  $i = 1, \dots, n$ . Com isso, o traço da matriz também é nulo,

$$\mathbf{A} = a_{11} + \dots + a_{nn} = 0 + \dots + 0 = 0, \quad \forall \mathbf{A} \in A.$$

A partir de subespaços vetoriais dados, podemos formar um novo subespaços, dado pela interseção entre os subespaços dados.

EXEMPLO 1.26 (Interseção de subespaços). Dados dois subespaços  $S_1, S_2$  de um mesmo espaço vetorial  $X$ , é fácil verificar que  $S_1 \cap S_2$  também é subespaço. Da mesma forma, se  $S_1, \dots, S_k$  são subespaços, onde  $k \in \mathbb{K}$ , então  $S_1 \cap \dots \cap S_k$  também é.

**1.5. Combinação linear.** Como vimos, a essência de um espaço vetorial são as operações de adição e multiplicação por escalar. Qual é o mínimo que podemos fazer com isso? *Combinações lineares!*

DEFINIÇÃO 1.5 (Combinação linear). *Seja  $X$  um espaço vetorial sobre um corpo  $\mathbb{K}$ . Dizemos que um elemento  $\mathbf{u} \in X$  é uma **combinação linear** de elementos  $\mathbf{v}_1, \dots, \mathbf{v}_n \in X$ ,  $n \in \mathbb{N}$ , quando*

$$\mathbf{u} = x_1 \mathbf{v}_1 + \dots + x_n \mathbf{v}_n,$$

para alguma combinação  $x_1, \dots, x_n \in \mathbb{K}$  de elementos do corpo.

EXEMPLO 1.27 (Combinação em espaços Euclidianos). O vetor  $\mathbf{u} = (1, 2)$  pode ser escrito como as seguintes combinações lineares

$$(2, 5) = 2(1, 0) + 5(0, 1),$$

$$(2, 5) = (2, 0) + (0, 5),$$

$$(2, 5) = 2(1, 1) + 3(0, 1),$$

$$(2, 5) = (1, 0) + (1, 1) + 2(0, 2),$$

$$(2, 5) = \frac{1}{2}(4, 10).$$

EXEMPLO 1.28 (Combinação de monômios). Um polinômio de grau  $n$  em  $\mathbb{R}$  é combinação linear de  $\{1, x, \dots, x^n\}$ .

EXEMPLO 1.29 (Combinação senoidal). Uma função senoidal  $x(t) = \cos(\theta + \omega t)$  com fase  $\theta$  arbitrária e frequência  $\omega$  dada pode ser escrita como combinação linear de um seno e um cosseno com mesma frequência e sem mudança de fase:

$$\cos(\theta + \omega t) = A_\theta \cos(\omega t) + B_\theta \sin(\omega t),$$

onde

$$A_\theta = \cos \theta, \quad B_\theta = -\sin \theta,$$

que segue da fórmula  $\cos(a + b) = \cos(a)\cos(b) - \sin(a)\sin(b)$ , com  $a = \theta$  e  $b = \omega t$ .

As operações de adição e multiplicação por escalar entre elementos do espaço vetorial  $X$  e do corpo  $\mathbb{K}$  associado pode ser estendida a subconjuntos de  $X$  e de  $\mathbb{K}$ :

DEFINIÇÃO 1.6 (Combinações lineares de conjuntos). *Seja  $X$  um espaço vetorial sobre um corpo  $\mathbb{K}$ . Dados conjuntos  $A, B \subset X$  e  $\Lambda \subset \mathbb{K}$ , além de elementos  $\mathbf{v} \in X$  e  $\lambda \in \mathbb{K}$ , podemos definir os conjuntos*

$$A + B = \{\mathbf{u} \in X; \exists \mathbf{a} \in A, \exists \mathbf{b} \in B; \mathbf{u} = \mathbf{a} + \mathbf{b}\},$$

$$\mathbf{v} + A = \{\mathbf{u} \in X; \exists \mathbf{a} \in A, \mathbf{u} = \mathbf{v} + \mathbf{a}\},$$

$$\Lambda A = \{\mathbf{u} \in X; \exists \mathbf{a} \in A, \exists \lambda \in \Lambda, \mathbf{u} = \lambda \mathbf{a}\},$$

$$\lambda A = \{\mathbf{u} \in X; \exists \mathbf{a} \in A, \mathbf{u} = \lambda \mathbf{a}\},$$

e

$$\Lambda \mathbf{v} = \{\mathbf{u} \in X; \exists \lambda \in \Lambda, \mathbf{u} = \lambda \mathbf{v}\}.$$

EXEMPLO 1.30 (Paralelogramo). Um paralelogramo  $\partial\Omega$  com vértices  $A, B, C, D$  no plano, onde o lado  $\overline{AB}$  é semelhante a  $\overline{CD}$  e o lado  $\overline{AC}$  é semelhante a  $\overline{BD}$  pode ser escrito como a seguinte soma de conjuntos, no espaço vetorial  $\mathbb{R}^2$ :

$$\partial\Omega = \{\overrightarrow{OA}, \overrightarrow{OB}, \overrightarrow{OC}, \overrightarrow{OD}\} = \overrightarrow{OA} + \{\overrightarrow{OO}, \overrightarrow{AB}\} + \{\overrightarrow{OO}, \overrightarrow{AC}\}$$

onde  $O$  representa a origem no plano, correspondendo a vetor nulo  $\mathbf{0} = \overrightarrow{OO}$  no espaço vetorial. Em relação à notação, observe que se  $P = (x, y)$  é um ponto no plano  $xy$  com coordenadas  $(x, y)$ , então interpretamos  $\overrightarrow{OP} = (x, y)$  como um vetor no espaço vetorial  $\mathbb{R}^2$ , enquanto que  $\overline{OP}$  indica o segmento de reta, no plano, que liga a origem do plano ao ponto  $P$ . Aqui, usamos  $\mathbb{R}^2$  para representar apenas o conjunto de pontos no plano  $xy$ , sem estrutura algébrica, como o espaço vetorial obtido a partir dele, juntando as operações de adição e multiplicação por escalar. Em alguns textos, há uma notação diferente para o espaço vetorial, como  $\mathbb{E}^n$ , reservando  $\mathbb{R}^n$  para o conjunto. Não vamos fazer essa distinção aqui, mas é importante saber que são dois objetos matemáticos diferentes.

EXEMPLO 1.31 (Interior de um paralelogramo). No exemplo [Exemplo 1.30](#), a região  $\Omega$  delimitada pelo paralelogramo  $\partial\Omega$  pode ser escrita como

$$\Omega = \overrightarrow{OA} + \{\overrightarrow{OO}, \overrightarrow{AB}\} + \{\lambda \overrightarrow{AC}; 0 \leq \lambda \leq 1\} = \overrightarrow{OA} + \{\overrightarrow{OO}, \overrightarrow{AB}\} + [0, 1]\{\overrightarrow{AC}\}$$

**1.6. Subespaço gerado e conjunto gerador.** O conjunto formado por todas as combinações lineares de determinados vetores tem a propriedade de qualquer combinação linear de seus vetores continuar no próprio conjunto. Ou seja, ele é um subespaço vetorial. Esse é um tipo de subespaço muito importante, que aparece de diversas formas e leva um nome especial.

**DEFINIÇÃO 1.7** (Subespaço gerado). *Seja  $X$  um espaço vetorial sobre um corpo  $\mathbb{K}$  e sejam  $\mathbf{v}_1, \dots, \mathbf{v}_k \in X$ ,  $k \in \mathbb{N}$ . O **subespaço gerado** por esses vetores é o conjunto formado por todas as combinações lineares desse vetores e é denotado por  $\text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ , i.e.*

$$\text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\} = \{\mathbf{v} \in X; \exists x_1, \dots, x_n \in \mathbb{K} \text{ tais que } \mathbf{v} = x_1\mathbf{v}_1 + \dots + x_n\mathbf{v}_n\}.$$

*Dizemos que  $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$  **gera** esse subespaço ou que é um **conjunto gerador** desse subespaço.*

*Mais geralmente, o subespaço gerado por um subconjunto  $A \subset X$  de  $X$  é o conjunto de todas as combinações lineares de  $A$ , i.e.*

$$\text{span}(A) = \{\mathbf{v} \in X; \exists n \in \mathbb{N}, \exists x_1, \dots, x_n \in \mathbb{K}, \\ \exists \mathbf{v}_1, \dots, \mathbf{v}_n \in A \text{ tais que } \mathbf{v} = x_1\mathbf{v}_1 + \dots + x_n\mathbf{v}_n\}.$$

**OBSERVAÇÃO 1.10.** Dado um subconjunto  $B \subset X$  de um espaço vetorial  $X$  e um subconjunto  $A$  tal que todo elemento de  $B$  pode ser escrito como combinação linear de elementos de  $A$ , isso não quer dizer que  $B = \text{span}(A)$ . O máximo que podemos afirmar nesse caso é a inclusão  $B \subset \text{span}(A)$ . Para que  $B = \text{span}(A)$ , precisamos ter que todas as combinações lineares de  $A$  estejam em  $B$ . Por exemplo, a bola unitária  $B = \{(x, y) \in \mathbb{R}^2; x^2 + y^2 \leq 1\}$  está contida no espaço gerado  $\text{span}\{(1, 0), (0, 1)\}$ , que é todo o  $\mathbb{R}^2$ .

**EXEMPLO 1.32** (geradores de  $\mathbb{R}^2$ ). Qualquer vetor  $(x, y) \in \mathbb{R}^2$  pode ser escrito em termos de um número finito de vetores, de várias formas:

$$\begin{aligned} (x, y) &= x(1, 0) + y(0, 1), \\ &= x(1, 1) + (y - x)(0, 1), \\ &= (x + y)(1, 1) - y(1, 0) - x(0, 1) \\ &= x(1, 1) + 0(1, 0) + (y - x)(0, 1) \\ &= y(1, 1) + (x - y)(1, 0) + 0(0, 1) \\ &= \dots, \end{aligned}$$

ou seja,

$$\begin{aligned}\mathbb{R}^2 &= \text{span}\{(1, 0), (0, 1)\} = \text{span}\{(1, 1), (0, 1)\} \\ &= \text{span}\{(1, 1), (1, 0), (0, 1)\} = \text{span}\{(2, 3), (3, 2)\}.\end{aligned}$$

EXEMPLO 1.33 (Senoidal). Fixado  $\omega \neq 0$ , temos, para  $A, \theta \in \mathbb{R}$  quaisquer,

$$A \cos(\theta + \omega t) = A \cos \theta \cos(\omega t) - A \sin \theta \sin(\omega t).$$

Além disso, dados  $a, b \in \mathbb{R}$ , definindo  $A$  por

$$A = \sqrt{a^2 + b^2},$$

vemos que

$$-1 \leq \frac{a}{A}, \frac{b}{A} \leq 1, \quad \left(\frac{a}{A}\right)^2 + \left(\frac{b}{A}\right)^2 = 1,$$

de maneira que existe um único  $\theta$  tal que

$$a = \cos \theta, \quad b = -\sin \theta.$$

Com isso,

$$a \cos(\omega t) + b \sin(\omega t) = A \cos \theta \cos(\omega t) - A \sin \theta \sin(\omega t) = A \cos(\theta + \omega t).$$

Ou seja,

$$S_\omega = \{A \cos(\theta + \omega t), A, \theta \in \mathbb{R}\} = \text{span}\{\cos(\omega t), \sin(\omega t)\}.$$

EXEMPLO 1.34 (Senoidal com amplitude fixa). No contexto do [Exemplo 1.33](#), se considerarmos apenas  $\omega$ , com amplitude constante  $A = 1$ , por exemplo, não temos o espaço gerado todo, ou seja, temos apenas a inclusão

$$\{\cos(\theta + \omega t), \theta \in \mathbb{R}\} \subset \text{span}\{\cos(\omega t), \sin(\omega t)\},$$

sendo que  $\{\cos(\theta + \omega t), \theta \in \mathbb{R}\}$  é subconjunto que não é um subespaço.

EXEMPLO 1.35 (Polinômios). Dado  $n$ , temos

$$\mathcal{P}_n(\mathbb{R}) = \{p(x) = a_0 + a_1x + \cdots + a_nx^n; a_0, \dots, a_n \in \mathbb{R}\} = \text{span}\{1, x, \dots, x^n\},$$

ao passo que, permitindo que  $n$  varie,

$$\mathcal{P}(\mathbb{R}) = \{p(x) = a_0 + a_1x + \cdots + a_nx^n; n \in \mathbb{N}, a_0, \dots, a_n \in \mathbb{R}\} = \text{span}(A),$$

onde

$$A = \{1, x, x^2, \dots\} = \bigcup_{n \in \mathbb{Z}, n \geq 0} \{x^n\}.$$

A combinação linear de subespaços vetoriais também dá um subespaço vetorial.

EXEMPLO 1.36 (Soma de subespaços). Dados dois subespaços  $S_1, S_2$  de um mesmo espaço vetorial  $X$ , é fácil verificar que  $S_1 + S_2$  também é subespaço. Da mesma forma, se  $S_1, \dots, S_k$  são subespaços, onde  $k \in \mathbb{K}$ , então  $S_1 + \dots + S_k$  também é.

## 2. Base e dimensão

### 2.1. Conjuntos linearmente dependentes e linearmente independentes.

Lembremos, agora, a definição de independência linear para um conjunto finito de vetores.

DEFINIÇÃO 2.1 (Conjuntos LI e LD). *Seja  $X$  um espaço vetorial sobre um corpo  $\mathbb{K}$ . Um conjunto finito de vetores  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\} \in X$  é dito **linearmente independente**, ou simplesmente **LI**, quando toda combinação linear satisfazendo*

$$x_1\mathbf{v}_1 + \dots + x_n\mathbf{v}_n = \mathbf{0},$$

onde  $x_1, \dots, x_n \in \mathbb{K}$ , é tal que

$$x_1 = \dots = x_n = 0.$$

Caso contrário, o conjunto é dito **linearmente dependente**, ou **LD**.

OBSERVAÇÃO 2.1. Observe que um conjunto formado por um único vetor não nulo  $\mathbf{u}$  é necessariamente linearmente independente, pois o único múltiplo dele que é igual ao vetor nulo é o múltiplo  $x\mathbf{u}$  obtido com o escalar  $x = 0$ . Observe, ainda, que um conjunto linearmente independente tem todos os seus vetores não nulos.

OBSERVAÇÃO 2.2. Conferir se um determinado conjunto finito de vetores em um espaço  $\mathbb{R}^m$  é linearmente independente ou não é equivalente à resolução de um sistema linear homogêneo. De fato, se  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\} \subset \mathbb{R}^m$ , escreva as coordenadas de cada vetor na forma  $\mathbf{v}_i = (v_{1i}, \dots, v_{mi})$ ,  $i = 1, \dots, n$ . Então a equação

$$x_1\mathbf{v}_1 + \dots + x_n\mathbf{v}_n = \mathbf{0}$$

tem a forma

$$x_1 \begin{pmatrix} v_{11} \\ \vdots \\ v_{m1} \end{pmatrix} + \dots + x_n \begin{pmatrix} v_{1n} \\ \vdots \\ v_{mn} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix},$$

que também pode ser escrito na forma matricial

$$\begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{m1} & \cdots & v_{mn} \end{bmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix},$$

que é equivalente ao sistema

$$\begin{cases} v_{11}x_1 + \cdots + v_{1n}x_n = 0, \\ \vdots \quad \ddots \quad \vdots \\ v_{m1}x_1 + \cdots + v_{mn}x_n = 0. \end{cases}$$

EXEMPLO 2.1 (Vetores canônicos em  $\mathbb{R}^n$ ). Os vetores  $\mathbf{e}_1 = (1, 0)$  e  $\mathbf{e}_2 = (0, 1)$  formam um conjunto linearmente independente em  $\mathbb{R}^2$ . De fato, o sistema associado à equação  $x_1\mathbf{e}_1 + x_2\mathbf{e}_2 = \mathbf{0}$  é simplesmente  $x_1 = 0$  e  $x_2 = 0$ . Mais geralmente, os vetores canônicos  $\mathbf{e}_1 = (1, 0, \dots, 0)$ ,  $\mathbf{e}_2 = (0, 1, 0, \dots, 0)$ ,  $\dots$ ,  $\mathbf{e}_n = (0, \dots, 0, 1)$  são linearmente independentes em  $\mathbb{R}^n$ , com o sistema homogêneo associado se tornando simplesmente  $x_j = 0$ ,  $j = 1, \dots, n$ . Qualquer subconjunto não-vazio de  $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$  também é linearmente independente.

EXEMPLO 2.2 (Funções linearmente independentes). Considere duas funções com valores reais  $f = f(x)$  e  $g = g(x)$  definidas em um intervalo  $I \subset \mathbb{R}$ . Podemos enxergá-las como vetores no espaço vetorial  $X = \mathbb{R}^I$ . O zero aditivo nesse espaço é a função identicamente nula, que também denotamos, com um abuso de notação, por 0. Suponha que, nesse espaço,  $\lambda f + \mu g = 0$ , para escalares  $\lambda, \mu \in \mathbb{R}$ , o que significa que

$$\lambda f(x) + \mu g(x) = 0,$$

para todo  $x$ . Se eles não forem linearmente independentes, então existem  $\lambda, \mu$  não ambos nulos tais que essa equação vale para todo  $x$ . Se apenas  $\mu = 0$ , então  $\lambda f(x) = 0$  para todo  $x$ , o que implica em  $f(x) = 0$  para todo  $x$ , ou seja,  $f = 0$ . Se apenas  $\lambda = 0$ , deduzimos, da mesma forma, que  $g = 0$ . Caso nenhuma das duas funções seja nula, devemos, então, ter ambos  $\lambda, \mu \neq 0$ , de modo que  $g(x) = -(\lambda/\mu)f(x)$ , ou seja, uma função é um múltiplo da outra e vice-versa. Em resumo, duas funções são linearmente dependentes caso pelo menos uma delas seja identicamente nula ou caso uma seja um múltiplo não-nulo da outra. É claro que se uma delas for nula, ela é um múltiplo da outra, com o múltiplo sendo o zero. Portanto, podemos concluir que *duas funções são linearmente dependentes se, e somente se, uma é um múltiplo da outra*.

**2.2. Subespaços linearmente independentes e soma direta.** Um combinação com propriedades especiais aparece quando os subespaços tem interseção trivial. Isso está relacionado à extensão do conceito de independência linear a subespaços.

PROPOSIÇÃO 2.1. *Se  $S_1$  e  $S_2$  são dois subespaços de um espaço vetorial  $X$  sobre um corpo  $\mathbb{K}$ , então  $S_1 \cap S_2 = \{\mathbf{0}\}$  se, e somente se, todo elemento de  $S_1 + S_2$  possui uma única representação como soma de um vetor em  $S_1$  e um vetor de  $S_2$ , i.e. para*

todo  $\mathbf{u} \in S_1 + S_2$ , existem únicos  $\mathbf{u}_1 \in S_1$  e  $\mathbf{u}_2 \in S_2$  tais que

$$\mathbf{u} = \mathbf{u}_1 + \mathbf{u}_2.$$

DEMONSTRAÇÃO. Suponha que  $S_1 \cap S_2 = \{0\}$ . Seja  $\mathbf{u} \in S_1 + S_2$ . Então, por definição da soma de subespaços, existem  $\mathbf{u}_1 \in S_1$  e  $\mathbf{u}_2 \in S_2$  tais que  $\mathbf{u} = \mathbf{u}_1 + \mathbf{u}_2$ . Para mostrar a unicidade, suponha que  $\mathbf{u} = \mathbf{v}_1 + \mathbf{v}_2$ , para outros dois vetores  $\mathbf{v}_1 \in S_1$  e  $\mathbf{v}_2 \in S_2$ . Então  $\mathbf{u} = \mathbf{u}_1 + \mathbf{u}_2 = \mathbf{v}_1 + \mathbf{v}_2$ . Subtraindo as duas representações, temos

$$\mathbf{u}_1 - \mathbf{v}_1 = \mathbf{u}_2 - \mathbf{v}_2,$$

com  $\mathbf{u}_1 - \mathbf{v}_1 \in S_1$  e  $\mathbf{u}_2 - \mathbf{v}_2 \in S_2$ . Logo,  $\mathbf{u}_1 - \mathbf{v}_1 = \mathbf{u}_2 - \mathbf{v}_2 \in S_1 \cap S_2$ . Como  $S_1 \cap S_2 = \{0\}$ , então

$$\mathbf{u}_1 - \mathbf{v}_1 = \mathbf{u}_2 - \mathbf{v}_2 = 0,$$

ou seja  $\mathbf{u}_1 = \mathbf{v}_1$  e  $\mathbf{u}_2 = \mathbf{v}_2$ , mostrando a unicidade. Agora suponha que valha a unicidade. Seja  $\mathbf{u} \in S_1 \cap S_2$ . Podemos escrever tanto  $\mathbf{u} = \mathbf{0} + \mathbf{u}$  com  $\mathbf{0} \in S_1$  e  $\mathbf{u} \in S_2$ , como  $\mathbf{u} = \mathbf{u} + \mathbf{0}$  com  $\mathbf{u} \in S_1$  e  $\mathbf{0} \in S_2$ . Como a representação é única, então devemos ter  $\mathbf{u} = \mathbf{0}$  tanto como elemento de  $S_1$  como como elemento de  $S_2$ , de forma que  $S_1 \cap S_2 = \{0\}$ .  $\square$

Essa ideia de representação em forma única como soma de elementos em diferentes subespaços vetoriais pode ser estendido a vários subespaços.

DEFINIÇÃO 2.2. Um subconjunto  $W \subset X$  de um subespaço vetorial  $X$  é dito soma direta de subespaços  $S_1, \dots, S_k$  quando todo elemento  $\mathbf{u}$  de  $W$  pode ser escrito de forma única como  $\mathbf{u} = \mathbf{u}_1 + \dots + \mathbf{u}_k$ , com  $\mathbf{u}_j \in S_j$ ,  $j = 1, \dots, k$ . Nesse caso, denotamos

$$W = S_1 \oplus S_2 \oplus \dots \oplus S_k.$$

Usando essa definição, podemos reescrever o resultado do [Proposição 2.1](#) da seguinte forma.

COROLÁRIO 2.1. Se  $S_1$  e  $S_2$  são dois subespaços de um espaço vetorial  $X$  sobre um corpo  $\mathbb{K}$ , então  $W \subset X$  é soma direta  $W = S_1 \oplus S_2$  se, e somente se,  $W = S_1 + S_2$  e  $S_1 \cap S_2 = \{0\}$ .

No caso de soma direta de vários subespaços, o critério é verificar que a única combinação que dá o vetor nulo é somando vetores nulos. Isso estende a condição  $S_1 \cap S_2 = \{0\}$  para vários subespaços, de forma adequado ao resultado de combinação linear única. Nesse sentido, temos o conceito de subespaços linearmente independentes.

DEFINIÇÃO 2.3 (Subespaços independentes). *Subespaços  $S_1, \dots, S_k \subset X$ ,  $k \in \mathbb{N}$ , de um espaço vetorial  $X$  são ditos **linearmente independentes**, ou simplesmente **independentes**, quando a única combinação linear de elementos em cada espaço que dá o vetor nulo é com vetores nulos, i.e. se*

$$\mathbf{u}_1 + \mathbf{u}_2 + \dots + \mathbf{u}_k = \mathbf{0},$$

com  $\mathbf{u}_j \in S_j$ ,  $j = 1, \dots, k$ , então

$$\mathbf{u}_1 = \mathbf{u}_2 = \dots = \mathbf{u}_k = \mathbf{0}.$$

Com isso, obtemos o seguinte resultado.

TEOREMA 2.1. *Um subconjunto  $W \subset X$  de um subespaço vetorial  $X$  é soma direta  $W = S_1 \oplus S_2 \oplus \dots \oplus S_k$  de subespaços  $S_1, \dots, S_k \subset X$ ,  $k \in \mathbb{N}$ , se, e somente se, os subespaços  $S_1, \dots, S_k$  são linearmente independentes entre si.*

DEMONSTRAÇÃO. A volta segue da unicidade da representação. Assumindo que  $W$  é soma direta dos subespaços  $S_j$  e observando que  $\mathbf{0} = \mathbf{u}_1 + \dots + \mathbf{u}_k$  para  $\mathbf{u}_j = \mathbf{0} \in S_j$ , segue da unicidade da representação que essa é a única adição possível de elementos dos subespaços  $S_j$  que dá o vetor nulo.

Suponha, agora, que essa seja a única adição possível que dá o vetor nulo. Vamos mostrar que as representações são únicas. Para isso, seja  $\mathbf{w} \in W$  e suponha que tenhamos duas representações,

$$\mathbf{w} = \mathbf{u}_1 + \mathbf{u}_2 + \dots + \mathbf{u}_k = \mathbf{v}_1 + \mathbf{v}_2 + \dots + \mathbf{v}_k,$$

com  $\mathbf{u}_j, \mathbf{v}_j \in S_j$ . Subtraindo as duas representações, obtemos

$$(\mathbf{u}_1 - \mathbf{v}_1) + (\mathbf{u}_2 - \mathbf{v}_2) + \dots + (\mathbf{u}_k - \mathbf{v}_k) = \mathbf{0}.$$

Como  $\mathbf{u}_j - \mathbf{v}_j \in S_j$ , visto que  $S_j$  é subespaço vetorial, e como a única adição que dá o vetor nulo é a de vetores nulos nos subespaços, então cada

$$\mathbf{u}_j - \mathbf{v}_j = \mathbf{0},$$

ou seja  $\mathbf{u}_j = \mathbf{v}_j$ , para cada  $j = 1, \dots, k$ , provando a unicidade da representação, o que significa que  $W = S_1 \oplus S_2 \oplus \dots \oplus S_k$  é soma direta dos subespaços  $S_1, S_2, \dots, S_k$ .  $\square$

**2.3. Base finita.** Essa ideia de representação única de vetores nos leva a uma definição clássica de base feita em dimensão finita, nos primeiros cursos de Álgebra Linear. Ressaltamos, aqui, que a ordem desse conjunto de vetores é fundamental, pois isso influi na representação. Portanto, a definição de base envolve uma *lista*, ou, dita de outra maneira, um conjunto *ordenado*.

DEFINIÇÃO 2.4 (Base finita). *Um conjunto ordenado finito de elementos de um espaço vetorial  $X$  é dito uma **base** de  $X$  quando todo vetor de  $X$  pode ser escrito de maneira única como combinação linear dos elementos do conjunto.*

OBSERVAÇÃO 2.3. Um conjunto ordenado, ou lista, costuma ser representada entre parênteses e intercalada por vírgulas, como no caso de um vetor em  $\mathbb{R}^n$ :

$$(x_1, \dots, x_n) \in \mathbb{R}^n.$$

Por sua vez, bases, assim como sequências, apesar de serem listas, costumam ser representadas entre chaves. Vamos manter essa convenção aqui para sequências, mas vamos usar uma notação diferente para conjuntos ordenados formando bases, através do uso de chaves duplas.

DEFINIÇÃO 2.5 (Notação para base). *Denotaremos bases genéricas através dos símbolos  $\mathbf{a}, \mathbf{b}, \mathbf{c}$  e bases canônicas através do símbolo  $\mathbf{e}$ . Se a base, digamos  $\mathbf{b}$ , é formada por  $n$  vetores  $\mathbf{u}_1, \dots, \mathbf{u}_n \in X$ ,  $n \in \mathbb{N}$ , usamos a notação de chaves duplas  $\{\{\cdot\}\}$ ,*

$$\mathbf{b} = \{\{\mathbf{u}_1, \dots, \mathbf{u}_n\}\}$$

para representar essa base.

OBSERVAÇÃO 2.4. Os vetores de uma base são necessariamente não nulos, caso contrário teríamos um infinidade de múltiplos do vetor nulo representando um mesmo vetor. Em particular, um espaço vetorial trivial não possui uma base.

OBSERVAÇÃO 2.5. Nem todo espaço vetorial possui uma base. Veja o espaço dos polinômios, por exemplo. Se pudéssemos escrever todos os polinômios como combinação linear de um conjunto finito  $\mathbf{b} = \{p_1, \dots, p_n\}$  de polinômios, estaríamos restritos a polinômios de grau no máximo igual ao grau do polinômio de maior grau dentre os polinômios do conjunto  $\mathbf{b}$ . Não temos como representar todos os polinômios desse jeito. Portanto, o espaço vetorial dos polinômios de grau arbitrário não possui uma base no sentido acima.

Uma caracterização importante de base, que muitas vezes é tomada como sendo a própria definição de base, é a de que os seus elementos formam um conjunto LI que gera todo o espaço. Colocamos isso, aqui, como resultado.

TEOREMA 2.2. *Um conjunto ordenado finito  $\mathbf{b} = \{\{\mathbf{v}_1, \dots, \mathbf{v}_n\}\}$  de elementos de um espaço vetorial  $X$  é uma base de  $X$  se, e somente se,  $\mathbf{b}$  é um conjunto linearmente independente e que gera todo o espaço  $X$ .*

DEMONSTRAÇÃO. Suponha que  $\mathcal{b} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  seja uma base de  $X$ . Se  $\mathcal{b}$  não fosse um conjunto linearmente independente, teríamos escalares  $x_1, \dots, x_n$ , nem todos nulos, tais que

$$x_1\mathbf{v}_1 + \dots + x_n\mathbf{v}_n = \mathbf{0}.$$

Por outro lado, também temos

$$0\mathbf{v}_1 + \dots + 0\mathbf{v}_n = \mathbf{0}.$$

Isso nos daria duas representações diferentes de  $\mathbf{0}$ , contrariando a nossa definição de base. Portanto,  $\mathcal{b}$  tem que ser LI. Por sua vez, se todo elemento de  $X$  pode ser escrito como combinação linear dos elementos de  $\mathcal{b}$ , então certamente  $X$  pertence ao espaço gerado por  $\mathcal{b}$ . Como o espaço gerado por  $\mathcal{b} \subset X$  deve ser, ele próprio, um subconjunto de  $X$ , então necessariamente  $X = \text{span}(\mathcal{b})$ .

Vamos supor agora que  $\mathcal{b} = \{\mathbf{v}_1, \dots, \mathbf{v}_2\}$  seja LI e gere  $X$ . Bom, isso já nos dá que todo elemento de  $X$  pode ser escrito como combinação linear de elementos de  $\mathcal{b}$ . Basta ver que isso deve ser feito de maneira única. Suponha, então, que tenhamos duas representações para um vetor  $\mathbf{v} \in X$ ,

$$\mathbf{v} = x_1\mathbf{v}_1 + \dots + x_n\mathbf{v}_n, \quad \mathbf{v} = y_1\mathbf{v}_1 + \dots + y_n\mathbf{v}_n.$$

Subtraindo as duas representações, temos

$$(x_1 - y_1)\mathbf{v}_1 + \dots + (x_n - y_n)\mathbf{v}_n = \mathbf{0}.$$

Como  $\mathcal{b}$  é LI, então todos os coeficientes acima devem se anular, o que significa que  $y_j = x_j$ , para todo  $j = 1, \dots, n$ , ou seja, a representação deve ser única. Isso mostra que  $\mathcal{b}$  é uma base, segundo a [Definição 2.4](#), completando a demonstração.  $\square$

**2.4. Dimensão.** Um ponto importante é o de que todas as bases de um mesmo espaço vetorial possuem o mesmo número de elementos. Isso nos permitir definir esse número como sendo a dimensão do espaço. Vejamos, primeiro, o seguinte resultado.

LEMA 2.1. *Seja  $X$  um espaço vetorial sobre um corpo  $\mathbb{K}$ . Considere dois conjuntos  $\mathcal{a} = \{\mathbf{u}_1, \dots, \mathbf{u}_k\}$  e  $\mathcal{b} = \{\mathbf{v}_1, \dots, \mathbf{v}_m\}$  de vetores em  $X$ , com  $k, m \in \mathbb{N}$ . Suponha que  $\mathcal{a}$  seja linearmente independente e que esteja contido no  $\text{span}$  de  $\mathcal{b}$ . Então  $k \leq m$ .*

DEMONSTRAÇÃO. Vamos provar por absurdo. Ou seja, assumir que  $k > m$  e chegar em uma contradição. Para isso, vamos usar indução. Antes de começar, note que todos os vetores de  $\mathcal{a}$  são não-nulos, visto que  $\mathcal{a}$  é linearmente independente.

Vamos mostrar que, a menos de uma reordenação dos vetores de  $\mathcal{b}$ , temos

$$\text{span}(\mathcal{b}) = \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_j, \mathbf{v}_{j+1}, \dots, \mathbf{v}_m\}, \quad j = 1, \dots, m.$$

Vamos começar com  $j = 1$ . Como  $\mathcal{a} \subset \text{span}(\mathcal{b})$ , podemos escrever cada vetor de  $\mathcal{a}$  como combinação linear de vetores em  $\mathcal{b}$ . Em particular, começando com  $\mathbf{u}_1$ , existem escalares  $a_{11}, \dots, a_{m1}$  tais que

$$\mathbf{u}_1 = a_{11}\mathbf{v}_1 + \dots + a_{m1}\mathbf{v}_m.$$

Como  $\mathbf{u}_1$  é não-nulo, existe pelo menos um coeficiente que é diferente de zero. Sem perda de generalidade (basta reordenar os vetores de  $\mathcal{b}$ ), vamos assumir que  $a_{11} \neq 0$ . Assim, podemos escrever  $\mathbf{v}_1$  em termos de  $\mathbf{u}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ , i.e.

$$\mathbf{v}_1 = \frac{1}{a_{11}} (\mathbf{u}_1 - a_{21}\mathbf{v}_2 - \dots - a_{m1}\mathbf{v}_m).$$

Ou seja,  $\mathbf{v}_1 \in \text{span}\{\mathbf{u}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ . Obviamente, para  $i = 2, \dots, m$ , o vetor  $\mathbf{v}_i$  também está nesse espaço gerado. Portanto,

$$\text{span}(\mathcal{b}) = \text{span}\{\mathbf{u}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}.$$

Agora, vamos assumir que o resultado seja válido para  $j - 1$ , com  $1 < j \leq m$ . Vamos mostrar que também vale para  $j + 1$ . Estamos assumindo, então, que

$$\text{span}(\mathcal{b}) = \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_{j-1}, \mathbf{v}_j, \mathbf{v}_{j+1}, \dots, \mathbf{v}_m\}.$$

Como  $\mathcal{a} \subset \text{span}(\mathcal{b}) = \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_{j-1}, \mathbf{v}_j, \mathbf{v}_{j+1}, \dots, \mathbf{v}_m\}$ , então todos os vetores de  $\mathcal{a}$  podem ser escritos como combinação linear dos vetores  $\mathbf{u}_1, \dots, \mathbf{u}_{j-1}, \mathbf{v}_j, \mathbf{v}_{j+1}, \dots, \mathbf{v}_m$ . Em particular, tomando agora  $\mathbf{u}_j \in \mathcal{a}$ , existem escalares  $a_{1j}, \dots, a_{mj}$  tais que

$$\mathbf{u}_j = a_{1j}\mathbf{u}_1 + \dots + a_{(j-1)j}\mathbf{u}_{j-1} + a_{jj}\mathbf{v}_j + a_{(j+1)j}\mathbf{v}_{j+1} + \dots + a_{mj}\mathbf{v}_m.$$

Como  $\mathbf{u}_1, \dots, \mathbf{u}_j$  também é LI, por ser um subconjunto do conjunto  $\mathcal{a}$ , que é LI, então não podemos ter todos os coeficientes seguintes,  $\lambda_{jj}, \dots, \lambda_{mj}$  nulos. Caso contrário escreveríamos  $\mathbf{u}_j$  em função de  $\mathbf{u}_1, \dots, \mathbf{u}_{j-1}$ . Portanto, algum dos coeficientes seguintes deve ser diferente de zero. Sem perda de generalidade (podemos reordenar os vetores restantes  $\mathbf{v}_j, \dots, \mathbf{v}_m$ , o que não interfere nos outros que já foram selecionados antes), podemos assumir que  $\lambda_{jj} \neq 0$ . Assim, escrevemos

$$\mathbf{v}_j = \frac{1}{a_{jj}} (\mathbf{u}_j - a_{1j}\mathbf{u}_1 - \dots - a_{(j-1)j}\mathbf{u}_{j-1} - a_{(j+1)j}\mathbf{v}_{j+1} - \dots - a_{mj}\mathbf{v}_m).$$

Com isso, obtemos que

$$\begin{aligned} \text{span}(\mathcal{b}) &= \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_{j-1}, \mathbf{v}_j, \mathbf{v}_{j+1}, \dots, \mathbf{v}_m\} \\ &= \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_{j-1}, \mathbf{u}_j, \mathbf{v}_{j+1}, \dots, \mathbf{v}_m\}, \end{aligned}$$

concluindo a indução.

Considerando, agora,  $j = m$ , a indução acima nos dá que

$$\text{span}(\mathcal{b}) = \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_m\}.$$

Mas como estamos assumindo  $k > m$ , podemos, agora, pegar  $\mathbf{u}_{m+1} \in \mathcal{a}$ . Como  $\mathbf{u}_{m+1} \in \mathcal{a} \subset \text{span}(\mathcal{b})$ , o resultado acima nos diz que  $\mathbf{u}_{m+1}$  é combinação linear de  $\mathbf{u}_1, \dots, \mathbf{u}_m$ . O que é impossível, pois estamos assumindo  $\mathcal{a}$  linearmente independente. Ou seja,  $k$  não pode ser maior do que  $m$ . Devemos, necessariamente, ter  $k \leq m$ .  $\square$

OBSERVAÇÃO 2.6. Observe, na demonstração do [Lema 2.1](#), que o processo de indução foi feito com a hipótese de que  $k > m$ , para se chegar em uma contradição, o que invalida, por sua vez, a própria hipótese de indução. No entanto, vale ressaltar que, no caso possível em que  $k \leq m$ , o processo de indução é válido, mas só podemos ir até  $j = k$ , ou seja, temos  $k \leq m$  e é verdade que, considerando conjuntos ordenados para relacionar com bases,

$$\text{span}(\mathcal{b}) = \text{span}\{\{\mathbf{u}_1, \dots, \mathbf{u}_j, \mathbf{v}_{\sigma(j+1)}, \dots, \mathbf{v}_{\sigma(m)}\}\}, \quad j = 1, \dots, k,$$

para alguma reordenação (permutação)  $\sigma : \{1, \dots, m\} \rightarrow \{1, \dots, m\}$  apropriada dos vetores de  $\mathcal{b}$ .

Agora podemos mostrar que duas bases de um mesmo espaço têm o mesmo número de elementos.

TEOREMA 2.3. *Duas bases de um mesmo espaço vetorial têm necessariamente o mesmo número de elementos.*

DEMONSTRAÇÃO. Sejam  $\mathcal{a} = \{\{\mathbf{v}_1, \dots, \mathbf{v}_n\}\}$  e  $\mathcal{b} = \{\{\mathbf{w}_1, \dots, \mathbf{w}_m\}\}$  duas bases de um mesmo espaço vetorial  $X$  sobre um corpo  $\mathbb{K}$ , com  $n, m \in \mathbb{N}$ . Queremos mostrar que  $n = m$ . Como  $\mathcal{a}$  é LI e  $\mathcal{b}$  gera  $X$ , então  $\mathcal{a}$  está contido no espaço gerado por  $\mathcal{b}$ . Segue, então, do [Lema 2.1](#), que  $n \leq m$ . Por outro lado,  $\mathcal{b}$  também é LI e está contido no espaço gerado por  $\mathcal{a}$ , o que implica em  $m \leq n$ . Logo,  $m = n$ .  $\square$

DEFINIÇÃO 2.6 (Dimensão finita). *Se um espaço vetorial  $X$  possui uma base composta de um número finito  $n \in \mathbb{N}$  de vetores, dizemos que  $X$  tem **dimensão finita**  $n$  e denotamos essa dimensão por  $\dim(X) = n$ , visto que esse número finito  $n$  independe da base escolhida, graças ao [Teorema 2.3](#). Caso  $X = \{\mathbf{0}\}$  seja o espaço trivial, dizemos que ele tem **dimensão nula**, i.e.  $\dim(X) = 0$ . Caso ele não seja trivial e nem tenha uma base finita, dizemos que  $X$  tem **dimensão infinita**.*

DEFINIÇÃO 2.7 (Base e dimensão de subespaço). *Como todo subespaço de um espaço vetorial é, em si, um espaço vetorial, as definições de base e dimensão se estendem da mesma forma para subespaços, ou seja, uma base de um subespaço  $S$  é um conjunto finito  $\sigma$  tal que todo elemento de  $S$  se escreve de forma única como*

elementos de  $\sigma$ . Nesse caso,  $S$  é de dimensão finita e  $\dim(S)$  é o número de vetores em  $\sigma$ . Caso  $S$  seja o subespaço trivial, temos  $\dim(S) = 0$ , também de dimensão finita.

EXEMPLO 2.3. Em  $\mathbb{R}^2$ , temos a **base canônica**  $e = \{\{\mathbf{e}_1, \mathbf{e}_2\}\}$ , onde  $\mathbf{e}_1 = (1, 0)$  e  $\mathbf{e}_2 = (0, 1)$ . Outras bases são  $\{(1, 1), (-1, 1)\}$ ,  $\{(1, 0), (1, 1)\}$ ,  $\{(2, 3), (3, 2)\}$  e assim por diante, uma infinidade de opções. Em  $\mathbb{R}^n$ , a base canônica é  $e = \{\{\mathbf{e}_1, \dots, \mathbf{e}_n\}\}$ , onde  $\mathbf{e}_j = (\delta_{ij})_{i=1, \dots, n}$ , e  $\delta_{ij}$  é o Delta de Kronecker. Quando trabalhando com mais de uma dimensão ao mesmo tempo, usamos índices para indicar a dimensão em questão, por exemplo, com  $e_n = \{\{\{\}\mathbf{e}_1^n, \dots, \mathbf{e}_n^n\}\}$  indicando a base de  $\mathbb{R}^n$ ,  $n \in \mathbb{N}$ .

EXEMPLO 2.4. Considere o conjunto de todas as soluções da equação diferencial

$$\frac{d^2x}{dt^2} = -x,$$

que é dado por

$$S = \{x \in \mathbb{R}^{\mathbb{R}}; x(t) = C_1 \cos t + C_2 \sin t, \forall t \in \mathbb{R}, C_1, C_2 \in \mathbb{R}\}.$$

Como  $\cos t$  e  $\sin t$  não são um o múltiplo do outro, essas duas funções são linearmente independentes (veja [Exemplo 2.2](#)). Assim, podemos escrever  $S = \text{span}\{\cos t, \sin t\}$ , de modo que  $e = \{\{\cos t, \sin t\}\}$  gera  $S$  e é um conjunto LI. Ou seja,  $e$  é uma base de  $S$  e  $\dim(S) = 2$ .

PROPOSIÇÃO 2.2. *Seja  $X$  um espaço vetorial de dimensão  $n$  e considere  $\mathfrak{b}_k = \{\{\mathbf{v}_1, \dots, \mathbf{v}_k\}\}$  um subconjunto ordenado linearmente independente em  $X$ , com  $k < n$ . Então  $\mathfrak{b}$  pode ser completado com  $n - k$  vetores  $\mathbf{v}_{k+1}, \dots, \mathbf{v}_n$  de forma que  $\mathfrak{b}' = \{\{\mathbf{v}_1, \dots, \mathbf{v}_n\}\}$  seja uma base de  $X$ .*

DEMONSTRAÇÃO. Como  $k < n$  e  $\mathfrak{b}_k$  é LI, então  $\mathfrak{b}_k$  não pode gerar  $X$ , caso contrário  $\mathfrak{b}_k$  seria uma base de  $X$  e  $X$  teria dimensão  $k$ , ao invés de  $n$ . Assim, existe algum vetor  $\mathbf{v}_{k+1} \in X$  que não está em  $\text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ . Como  $\mathbf{v}_{k+1}$  não está nesse espaço gerado, então o conjunto ordenado  $\mathfrak{b}_{k+1} = \{\{\mathbf{v}_1, \dots, \mathbf{v}_{k+1}\}\}$  tem que ser linearmente independente. Prosseguimos assim  $n - k$  vezes, até chegar em  $\mathfrak{b}_n = \{\{\mathbf{v}_1, \dots, \mathbf{v}_n\}\}$  linearmente independente. Nesse caso,  $\mathfrak{b}_n$  tem que ser uma base de  $X$ , caso contrário poderíamos achar um outro vetor e obter um conjunto linearmente independente com  $n + 1$  vetores. Mas isso violaria o [Lema 2.1](#), já que  $X$  possui uma base (em particular um conjunto gerador) com  $n$  elementos. Concluimos, então, que  $\mathfrak{b}_n$  é uma base de  $X$  que completa os vetores em  $\mathfrak{b}_k$ .  $\square$

PROPOSIÇÃO 2.3. *Seja  $X$  um espaço vetorial de dimensão finita. Se  $S \subset X$  é um subespaço não trivial de  $X$ , então  $S$  possui uma base e  $\dim(S) \leq \dim(X)$ .*

DEMONSTRAÇÃO. Basta seguir um raciocínio semelhante ao da demonstração da **Proposição 2.2**, construindo conjuntos linearmente independentes em  $S$  até completar a dimensão de  $S$ . Como  $X$  tem dimensão finita, esse argumento necessariamente termina em um número finito de passos.  $\square$

Quando dois subespaços não têm interseção trivial, a dimensão do subespaço gerado por eles é menor do que a soma das dimensões desses subespaços. Mais precisamente, temos o seguinte resultado, cuja demonstração é deixada como exercício.

PROPOSIÇÃO 2.4. *Sejam  $U$  e  $V$  subespaços vetoriais de dimensão finita de um certo espaço vetorial. Então*

$$\dim(U + V) = \dim(U) + \dim(V) - \dim(U \cap V).$$

DEMONSTRAÇÃO. Quando  $U \cap V = \{\mathbf{0}\}$  é trivial, temos uma soma direta  $U + V = U \oplus V$  e  $\dim(U \cap V) = 0$ , de modo que a identidade vale. Quando  $U \cap V$  não é trivial, a ideia é partir de uma base de  $U \cap V$  e, por um lado, completar essa base até uma base de  $U$  e, por outro, completar até uma base de  $V$ . Juntando os elementos dessas bases, obtemos uma base para  $U + V$ . Em seguida, é só contabilizar os elementos das bases de  $U \cap V$ , de  $U$  e de  $V$ . Deixamos isso como exercício.  $\square$

A combinação de bases de subespaços linearmente independentes será fundamental em vários momentos. Em relação a isso, fazemos a seguinte definição.

DEFINIÇÃO 2.8. *Ao concatenarmos bases de subespaços linearmente independentes, digamos*

$$\mathfrak{a} = \{\mathbf{u}_1, \dots, \mathbf{u}_k\}, \quad \mathfrak{b}_2 = \{\mathbf{v}_1, \dots, \mathbf{v}_m\},$$

*bases de subespaços linearmente independentes  $U$  e  $V$ , respectivamente, a base de  $S = U \oplus V$  pode ser denotada por*

$$\mathfrak{c} = \{\{\mathfrak{a}, \mathfrak{b}\}\} = \{\{\{\mathbf{u}_1, \dots, \mathbf{u}_k\}, \{\mathbf{v}_1, \dots, \mathbf{v}_m\}\}\} = \{\{\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{v}_1, \dots, \mathbf{v}_m\}\}.$$

*Outra base possível é*

$$\tilde{\mathfrak{c}} = \{\{\mathfrak{b}, \mathfrak{a}\}\} = \{\{\mathbf{v}_1, \dots, \mathbf{v}_m, \mathbf{u}_1, \dots, \mathbf{u}_k\}\}.$$

*A ordem para a base é fundamental. Mas observe que a ordem para a definição do subespaço  $S = U \oplus V = V \oplus U$  é irrelevante. No caso de vários subespaços linearmente independentes  $S_1, \dots, S_l$ ,  $l \in \mathbb{N}$ , com bases  $\mathfrak{b}_1, \dots, \mathfrak{b}_l$ , denotamos a concatenação das bases, como base de  $S = S_1 \oplus \dots \oplus S_l$ , por*

$$\mathfrak{b} = \{\{\mathfrak{b}_1, \dots, \mathfrak{b}_l\}\}.$$

EXEMPLO 2.5. O espaço  $\mathbb{R}^3$  pode ser escrito por

$$\mathbb{R}^3 = U \oplus V,$$

onde

$$U = \{(x, y, 0); x, y \in \mathbb{R}\}, \quad V = \{(0, 0, z); z \in \mathbb{R}\}.$$

Considerando as bases

$$\mathfrak{a} = \{(1, 1, 0), (-1, 1, 0)\}, \quad \mathfrak{b} = \{(0, 0, 1)\}$$

de  $U$  e  $V$ , respectivamente, obtemos uma base de  $\mathbb{R}^3$ ,

$$\mathfrak{c} = \{\{\mathfrak{a}, \mathfrak{b}\}\} = \{(1, 1, 0), (-1, 1, 0), (0, 0, 1)\}.$$

Agora, pensando em objetos típicos da geometria Euclidiana, fazemos as seguintes definições.

DEFINIÇÃO 2.9 (Retas e planos). *Um subespaço de dimensão um de um espaço vetorial  $X$  é dito uma **reta** passando pela origem. Um subespaço de dimensão dois é dito um **plano** passando pela origem.*

Mais geralmente, temos

DEFINIÇÃO 2.10 (Retas, planos e subespaços afim). *Um **subespaço afim** de um espaço vetorial  $X$  é um subconjunto de  $X$  da forma  $\mathbf{u} + S$ , onde  $\mathbf{u} \in X$  e  $S$  é um subespaço de  $X$ . Caso o subespaço afim tenha dimensão um, dizemos que o subespaço afim é uma **reta afim**. Caso o subespaço tenha dimensão dois, dizemos que é um **plano afim**. Caso o subespaço  $S$  tenha co-dimensão um (veja [Seção 2.6](#) para o conceito de co-dimensão), dizemos que é um **hiperplano afim**.*

DEFINIÇÃO 2.11 (Subespaços afim paralelos). *Dois subespaços afim  $\mathbf{u} + S$  e  $\mathbf{u}' + S'$  são ditos **paralelos** quando  $S = S'$ .*

DEFINIÇÃO 2.12 (Segmento de reta). *Um **segmento de reta** em um espaço vetorial  $X$  sobre  $\mathbb{R}$  é um conjunto da forma  $L = \{(1 - \theta)\mathbf{u} + \theta\mathbf{v}; 0 \leq \theta \leq 1\}$ , onde  $\mathbf{u}, \mathbf{v} \in X$ . Os vetores  $\mathbf{u}$  e  $\mathbf{v}$  são os pontos **extremos** do segmento de reta.*

DEFINIÇÃO 2.13 (Cone). *Um **cone** em um espaço vetorial  $X$  sobre  $\mathbb{R}$  é um conjunto  $C$  com a propriedade de que  $\lambda\mathbf{v} \in C$  para todo  $\lambda > 0$  e todo  $\mathbf{v} \in C$ .*

DEFINIÇÃO 2.14 (Convexo). *Um conjunto  $C$  em um espaço vetorial  $X$  sobre  $\mathbb{R}$  é dito **convexo** quando, para todo  $\mathbf{u}, \mathbf{v} \in C$ , temos  $(1 - \theta)\mathbf{u} + \theta\mathbf{v} \in C$ , para todo  $0 \leq \theta \leq 1$ .*

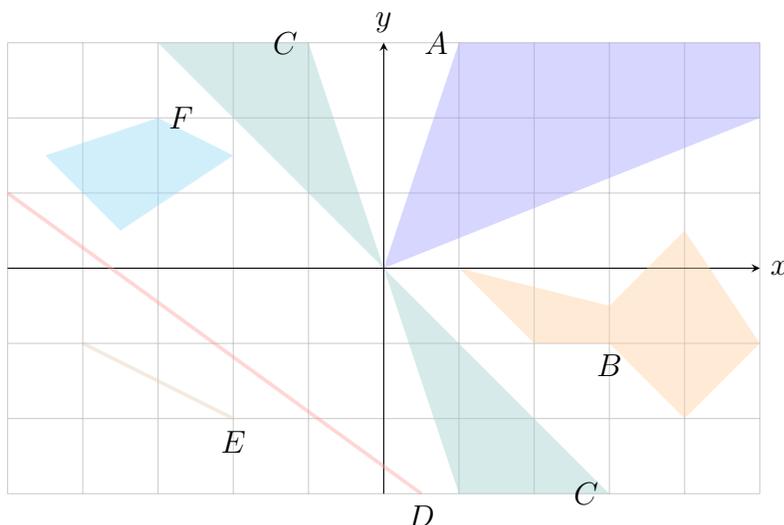


FIGURA 2.1. Tipos de subconjuntos de um espaço vetorial ( $\mathbb{R}^2$ ): cone convexo (A); formato estrela (B); cone não convexo (C); reta afim (D); segmento de reta (E) e convexo (F).

DEFINIÇÃO 2.15 (Formato estrela). Um conjunto  $C$  em um espaço vetorial  $X$  sobre  $\mathbb{R}$  é dito de **formato estrela** quando existe um vetor  $\mathbf{u} \in C$  tal que para todo vetor  $\mathbf{v} \in C$ , temos  $(1 - \theta)\mathbf{u} + \theta\mathbf{v} \in C$ , para todo  $0 \leq \theta \leq 1$ .

DEFINIÇÃO 2.16 (Simétrico). Um conjunto  $A$  em um espaço vetorial  $X$  é dito **simétrico** quando  $-\mathbf{u} \in A$  para todo  $\mathbf{u} \in A$ .

**2.5. Representação vetorial em uma base.** A definição de base nos leva a uma representação de um elemento de um espaço vetorial como um vetor em  $\mathbb{K}^n$ , onde  $\mathbb{K}$  é o corpo associado ao espaço vetorial e  $n$  é a dimensão do mesmo.

DEFINIÇÃO 2.17 (Representação vetorial em uma base). Seja  $X$  um espaço vetorial de dimensão finita  $n = \dim(X)$  sobre um corpo  $\mathbb{K}$  e seja  $\mathcal{b} = \{\{\mathbf{w}_1, \dots, \mathbf{w}_n\}\}$  uma base de  $X$ . Dado  $\mathbf{u} \in X$ , a **representação de  $\mathbf{u}$  na base  $\mathcal{b}$**  é o vetor  $(\mathbf{u})_{\mathcal{b}} \in \mathbb{K}^n$  dado por

$$(\mathbf{u})_{\mathcal{b}} = (x_1, \dots, x_n),$$

onde  $x_1, \dots, x_n \in \mathbb{K}$  são os escalares definidos unicamente pela combinação linear de  $\mathbf{u}$  em termos dos elementos da base,

$$\mathbf{u} = x_1\mathbf{w}_1 + \dots + x_n\mathbf{w}_n.$$

A representação gera um mapeamento  $\mathbf{u} \mapsto (\mathbf{u})_{\mathfrak{b}}$  que é, de fato, um isomorfismo entre  $X$  e  $\mathbb{K}^n$ .

**TEOREMA 2.4.** *Seja  $X$  um espaço vetorial de dimensão finita  $n = \dim(X)$  sobre um corpo  $\mathbb{K}$  e seja  $\mathfrak{b} = \{\{\mathbf{w}_1, \dots, \mathbf{w}_n\}\}$  uma base de  $X$ . Então a função  $J : V \rightarrow \mathbb{K}^n$  dada pela representação  $\mathbf{u} \mapsto J(\mathbf{u}) = (\mathbf{u})_{\mathfrak{b}}$  de  $\mathbf{u}$  na base  $\mathfrak{b}$  é um isomorfismo entre  $X$  e  $\mathbb{K}^n$ .*

**DEMONSTRAÇÃO.** A demonstração segue diretamente da linearidade do espaço e do fato da representação de um vetor como combinação linear da base ser única. De fato, se

$$\mathbf{u} = x_1 \mathbf{w}_1 + \dots + x_n \mathbf{w}_n$$

e

$$\mathbf{v} = y_1 \mathbf{w}_1 + \dots + y_n \mathbf{w}_n,$$

então

$$\begin{aligned} \mathbf{u} + \lambda \mathbf{v} &= x_1 \mathbf{w}_1 + \dots + x_n \mathbf{w}_n + \lambda(y_1 \mathbf{w}_1 + \dots + y_n \mathbf{w}_n) \\ &= (x_1 + \lambda y_1) \mathbf{w}_1 + \dots + (x_n + \lambda y_n) \mathbf{w}_n. \end{aligned}$$

Como a representação é única, necessariamente

$$J(\mathbf{u} + \lambda \mathbf{v}) = (\mathbf{u} + \lambda \mathbf{v})_{\mathfrak{b}} = (x_1 + \lambda y_1, \dots, x_n + \lambda y_n),$$

de modo que

$$J(\mathbf{u} + \lambda \mathbf{v}) = (\mathbf{u})_{\mathfrak{b}} + \lambda(\mathbf{v})_{\mathfrak{b}} = J(\mathbf{u}) + \lambda J(\mathbf{v}),$$

provando a linearidade de  $J$ . Para ver que  $J$  é sobrejetivo, basta observar que qualquer  $(x_1, \dots, x_n) \in \mathbb{K}^n$  é a imagem do elemento

$$\mathbf{u} = x_1 \mathbf{w}_1 + \dots + x_n \mathbf{w}_n.$$

□

Exploraremos essa representação vetorial mais adiante. No momento, vejamos, apenas, alguns exemplos.

**EXEMPLO 2.6.** Tomando a base canônica  $e = \{\{\mathbf{e}_1, \mathbf{e}_2\}\}$  de  $\mathbb{R}^2$ , onde  $\mathbf{e}_1 = (1, 0)$  e  $\mathbf{e}_2 = (0, 1)$ , podemos escrever qualquer  $\mathbf{u} = (x, y) \in \mathbb{R}^2$  como  $\mathbf{u} = x\mathbf{e}_1 + y\mathbf{e}_2$ , de modo que  $(\mathbf{u})_e = (x, y)$  é o próprio vetor. Em outra base, digamos  $\mathfrak{b} = \{\{\mathbf{w}_1, \mathbf{w}_2\}\}$  com  $\mathbf{w}_1 = (1, 1)$  e  $\mathbf{w}_2 = (-1, 1)$ , temos

$$\mathbf{u} = (x, y) = \left( \frac{x+y}{2} + \frac{x-y}{2}, \frac{x+y}{2} - \frac{x-y}{2} \right) = \frac{x+y}{2} \mathbf{w}_1 - \frac{x-y}{2} \mathbf{w}_2,$$

de modo que

$$(\mathbf{u})_e = \left( \frac{x+y}{2}, -\frac{x-y}{2} \right).$$

EXEMPLO 2.7. Na base  $e = \{1, x, x^2\}$ , um polinômio  $p(x) = a_0 + a_1x + a_2x^2$  em  $\mathcal{P}_2(\mathbb{R})$  é representado pelo vetor

$$(p)_e = (a_0, a_1, a_2) \in \mathbb{R}^3.$$

Em particular,

$$(1 + 2x^2)_{\{1, x, x^2\}} = (1, 0, 2).$$

EXEMPLO 2.8. A base canônica do espaço  $\mathbb{R}^{2 \times 2}$  das matrizes reais  $2 \times 2$  é

$$e = \left\{ \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\} \right\}.$$

Nesse espaço, temos

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + d \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

ou seja,

$$(\mathbf{A})_e = (a, b, c, d) \in \mathbb{R}^4.$$

EXEMPLO 2.9. Considere, agora, uma matriz simétrica em  $\mathbb{R}^{2 \times 2}$ , i.e. da forma

$$\mathbf{A} = \begin{bmatrix} a & b \\ b & d \end{bmatrix}.$$

Na base canônica de  $\mathbb{R}^{2 \times 2}$ , temos

$$(\mathbf{A})_e = (a, b, b, d) \in \mathbb{R}^4.$$

Uma base para esse subespaço de matrizes simétricas é

$$e = \left\{ \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\} \right\}.$$

Nesse subespaço, temos

$$\mathbf{A} = \begin{bmatrix} a & b \\ b & d \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + d \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix},$$

ou seja,

$$(\mathbf{A})_e = (a, b, d) \in \mathbb{R}^3.$$

Por exemplo,

$$\left( \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \right)_e = (1, 2, 5).$$

Essa estrutura simplificada é explorada em linguagens de programação, economizando o espaço de memória necessário para alocar a matriz e acelerando diversas operações matriciais envolvendo essas e outras matrizes especiais. Veja [Seção 4.2](#).

EXEMPLO 2.10 (Matrizes diagonais). Uma classe particular de matrizes simétricas é a de matrizes quadradas diagonais  $\mathbf{D} = (d_{ij})_{ij=1}^n$ , onde  $d_{ij} = 0$ , quando  $i \neq j$ . A multiplicação por escalar e a adição de matrizes diagonais continua sendo uma matriz diagonal, de forma que esse conjunto de matrizes diagonais forma um subespaço do espaço de matrizes quadradas. Uma base para esse espaço é formado pelas matrizes  $\mathbf{D}_k = (d_{ij}^{(k)})_{ij=1}^n$  com  $d_{ij}^{(k)} = 1$ , quando  $i = j = k$  e  $d_{ij}^{(k)} = 0$ , caso contrário. Em  $\mathbb{R}^{2 \times 2}$ , por exemplo, temos

$$\mathbf{D} = \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix} = d_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + d_2 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = d_1 D_1 + d_2 D_2.$$

EXEMPLO 2.11. O subespaço de  $X = \mathcal{C}(\mathbb{R})$  das soluções da equação diferencial de segunda ordem

$$\frac{d^2 y}{dx^2} - 4 \frac{dy}{dx} + 4y = 0$$

é determinado pela solução geral

$$U = \{y \in \mathcal{C}(\mathbb{R}); y(x) = (C_1 + C_2 x)e^{2x}, C_1, C_2 \in \mathbb{R}\}.$$

Uma base natural para esse espaço é dada por  $\mathfrak{b} = \{e^{2x}, xe^{2x}\}$ . Por exemplo, a solução particular  $y(x) = (1 + 2x)e^x$  tem a representação

$$(y)_{\mathfrak{b}} = (1, 2).$$

OBSERVAÇÃO 2.7. Em Geometria Diferencial, assim como em suas aplicações em Física Matemática, os vetores são chamados de *contra-variantes*. Isso por conta de como os coeficientes de um elemento do espaço vetorial variam conforme mudanças de escala nos elementos da base. Mais precisamente, suponha que  $\mathfrak{b} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  seja uma base de um espaço vetorial  $X$  de dimensão  $n \in \mathbb{N}$  e seja

$$\mathbf{u} = x_1 \mathbf{w}_1 + \dots + x_n \mathbf{w}_n$$

um elemento do espaço, com coeficientes  $x_i \in \mathbb{K}$  nessa base. Considere, então, uma nova base  $\tilde{\mathfrak{b}} = \{\tilde{\mathbf{w}}_1, \dots, \tilde{\mathbf{w}}_n\}$ , onde  $\tilde{\mathbf{w}}_i = \lambda \mathbf{w}_i$ , para algum  $\lambda > 0$ , representando uma mudança de escala na base. Nessa nova base, temos

$$\mathbf{u} = \frac{x_1}{\lambda} \lambda \mathbf{w}_1 + \dots + \frac{x_n}{\lambda} \lambda \mathbf{w}_n = \frac{x_1}{\lambda} \tilde{\mathbf{w}}_1 + \dots + \frac{x_n}{\lambda} \tilde{\mathbf{w}}_n.$$

Ou seja, enquanto a base mudou de acordo com

$$\mathbf{w}_i \mapsto \lambda \mathbf{w}_i$$

os coeficientes mudaram de forma inversa,

$$x_i \mapsto \frac{x_i}{\lambda}.$$

Por isso os vetores são chamados de contra-variantes. Isso aparece em Geometria Diferencial nos espaços tangentes de uma variedade, em relação a mudanças de coordenadas na variedade. Compare com a **Observação 3.3** que comenta sobre os funcionais lineares (ou co-vetores) serem *co-variantes*.

## 2.6. Espaço complementar e co-dimensão.

**PROPOSIÇÃO 2.5** (Complemento de um subespaço). *Seja  $S$  um subespaço de um espaço vetorial  $X$  de dimensão finita. Então existe um subespaço  $S'$  de  $X$  tal que  $X$  é soma direta de  $S$  por  $S'$ , ou seja,*

$$X = S \oplus S'.$$

**DEMONSTRAÇÃO.** Se  $S = \{\mathbf{0}\}$ , basta tomar  $S' = X$ , de modo que todo  $\mathbf{u} \in X = S \oplus S'$  se escreve de forma única  $\mathbf{u} = \mathbf{0} + \mathbf{u}$ , com  $\mathbf{0} \in S$  e  $\mathbf{u} \in S' = X$ .

Se  $S$  não é trivial, então segue da **Proposição 2.3** que  $S$  possui uma base, digamos  $\mathcal{a} = \{\{\mathbf{v}_1, \dots, \mathbf{v}_k\}\}$ , onde  $k = \dim(S) \leq \dim(X)$ . Se  $k = \dim(S) = \dim(X)$ , então  $S = X$  e basta tomarmos  $S' = \{\mathbf{0}\}$ . Nesse caso, todo  $\mathbf{u} \in X = S$  se escreve de maneira única como  $\mathbf{u} = \mathbf{u} + \mathbf{0}$ , com  $\mathbf{0} \in S'$ .

Caso  $k = \dim(S) < \dim(X) = n$ , então, graças ao **Proposição 2.2**, existem vetores  $\mathbf{v}_{k+1}, \dots, \mathbf{v}_n$  em  $V \setminus S$  tais que  $\mathcal{b} = \{\{\mathbf{v}_1, \dots, \mathbf{v}_n\}\}$  é uma base de  $X$ . Tomando  $S' = \text{span}\{\mathbf{v}_{k+1}, \dots, \mathbf{v}_n\}$ , vemos que  $S'$  é um subespaço de  $X$ , com dimensão  $n - k = \dim(X) - \dim(S)$ , tal que  $S + S' = X$  e ainda  $S \cap S' = \{\mathbf{0}\}$ . Ou seja,  $X = S \oplus S'$  é soma direta de  $S$  com  $S'$ .  $\square$

Uma consequência da demonstração é que  $\dim(S') = \dim(X) - \dim(S)$ , independente de como a base de  $S$  foi completada até uma base de  $X$ . Ou seja, temos o seguinte resultado.

**PROPOSIÇÃO 2.6.** *Seja  $S$  um subespaço próprio de um espaço vetorial  $X$  de dimensão finita. Então todo subespaço  $S'$  complementar a  $S$  possui a mesma dimensão  $\dim(S') = \dim(X) - \dim(S)$ .*

Deixamos os detalhes da demonstração desse fato a cargo do leitor. Com base nisso, podemos fazer a seguinte definição.

**DEFINIÇÃO 2.18** (Co-dimensão). *A **co-dimensão** de um subespaço de um espaço vetorial é a dimensão de um subespaço complementar.*

DEFINIÇÃO 2.19 (Hiperplanos e hiperplanos afim). *Um subespaço de co-dimensão um de um espaço vetorial  $X$  é dito uma **hiperplano**. Um subespaço afim  $\mathbf{u} + S$  onde  $\mathbf{u} \in X$  e  $S$  é um subespaço de co-dimensão um é dito um **hiperplano afim**.*

**2.7. Espaço quociente a um subespaço.** Dado um subespaço vetorial  $S$  de um espaço vetorial  $X$  sobre um corpo  $\mathbb{K}$ , temos associado a ele uma relação de equivalência entre vetores  $\mathbf{u}, \mathbf{v} \in X$  dada por

$$\mathbf{u} \sim \mathbf{v} \iff \mathbf{u} - \mathbf{v} \in S.$$

Essa relação de equivalência também é denotada na forma

$$\mathbf{u} \equiv \mathbf{v} \pmod{S},$$

indicando, mais explicitamente, que  $\mathbf{u}$  e  $\mathbf{v}$  são congruentes módulo  $S$ .

O fato disso ser uma relação de equivalência nos permite decompor o espaço na união disjunta das classes de equivalência. Uma **classe de equivalência** é um conjunto  $A \subset X$  tal que

- (1)  $A$  é não vazio;
- (2) Se  $\mathbf{u}, \mathbf{v} \in A$ , então  $\mathbf{u} \sim \mathbf{v}$ , ou seja, todos os elementos da classe são equivalentes entre si; e
- (3) Se  $\mathbf{u} \in A$  e  $\mathbf{v} \in X$  são tais que  $\mathbf{u} \sim \mathbf{v}$ , então  $\mathbf{v} \in A$ , ou seja,  $A$  contém todos os elementos equivalentes da classe.

Denotamos uma classe de equivalência pelo símbolo  $[[\cdot]]$ . Mais precisamente, dado um certo elemento  $\mathbf{u} \in X$ , denotamos o conjunto de todos os elementos equivalentes a  $\mathbf{u}$  por

$$[[\mathbf{u}]] = \{\mathbf{v} \in X; \mathbf{v} \sim \mathbf{u}\}.$$

Observe que podemos escrever

$$[[\mathbf{u}]] = \{\mathbf{v}; \mathbf{v} \sim \mathbf{u}\} = \{\mathbf{v}; \mathbf{w} = \mathbf{v} - \mathbf{u} \in S\} = \{\mathbf{u} + \mathbf{w}; \mathbf{w} \in S\} = \mathbf{u} + S.$$

Um fato fundamental de relações de equivalência é o de que duas classes de equivalência ou são disjuntas ou são iguais. Por conta disso, para quaisquer dois vetores  $\mathbf{u}$  e  $\mathbf{v}$  na mesma classe de equivalência, i.e.  $\mathbf{u} \sim \mathbf{v}$ , temos  $\mathbf{u} - \mathbf{v} \in S$ , de modo que

$$[[\mathbf{u}]] = \mathbf{u} + S = \mathbf{v} + (\mathbf{u} - \mathbf{v}) + S = \mathbf{v} + S = [[\mathbf{v}]],$$

Isso nos permite operar nas classes de equivalência através de seus representantes, sem que certas operações dependam do representante: podemos somar duas classes de equivalência  $[[\mathbf{u}_1]]$  e  $[[\mathbf{u}_2]]$  para obter a classe de equivalência  $[[\mathbf{u}_1 + \mathbf{u}_2]]$ ; e podemos multiplicar uma das classes por  $\lambda$  para obter  $[[\lambda\mathbf{u}]]$ .

Assim, temos um espaço quociente, denotado por  $X/S$ , formado pelo conjunto das classes de equivalência, de tal forma que as operações de adição de vetores e multiplicação por escalar em  $X$  se estendem naturalmente a  $X/S$ :

DEFINIÇÃO 2.20 (Espaço quociente). *Seja  $X$  um espaço vetorial sobre um corpo  $\mathbb{K}$  e seja  $S \subset X$  um subespaço vetorial de  $X$ . Então o espaço quociente de  $X$  por  $S$ , denotado por  $X/S$ , é o conjunto das classes de equivalência de  $\sim$ , que é um conjunto de subconjuntos de  $X$  dado por*

$$X/S = \{W \subset X; \exists \mathbf{u} \in X, W = [\mathbf{u}]\} = \{[\mathbf{u}]; \mathbf{u} \in X\} = \{\mathbf{u} + S; \mathbf{u} \in X\},$$

munido das operações de adição e multiplicação por escalar definidas por

$$[\mathbf{u}_1] \oplus [\mathbf{u}_2] = [\mathbf{u}_1 + \mathbf{u}_2]$$

e

$$\lambda \odot [\mathbf{u}] = [\lambda \mathbf{u}].$$

Levando em consideração a representação  $[\mathbf{u}] = \mathbf{u} + S$ , a adição em  $X/S$  pode ser escrita como

$$[\mathbf{u}_1] \oplus [\mathbf{u}_2] = (\mathbf{u}_1 + S) \oplus (\mathbf{u}_2 + S) = \mathbf{u}_1 + \mathbf{u}_2 + S$$

e a multiplicação como

$$\lambda \odot [\mathbf{u}] = \lambda \odot (\mathbf{u} + S) = \lambda \mathbf{u} + S.$$

Observe que, quando  $\lambda \neq 0$ , então podemos escrever a multiplicação por escalar em  $X/S$  como multiplicação em conjuntos:

$$\lambda \odot [\mathbf{u}] = \lambda \mathbf{u} + S = \lambda \mathbf{u} + \lambda S = \lambda(\mathbf{u} + S) = \lambda [\mathbf{u}],$$

mas quando  $\lambda = 0$ , isso não é possível, caso contrário o resultado seria o espaço trivial  $\{0\}$  ao invés do subespaço  $S$ .

PROPOSIÇÃO 2.7. *A co-dimensão de um subespaço  $S$  de um espaço vetorial  $X$  de dimensão finita é igual a dimensão do espaço vetorial quociente  $X/S$ , de modo que*

$$\dim(X) = \dim(S) + \dim(X/S). \quad (2.1)$$

DEMONSTRAÇÃO. Se  $X = \{\mathbf{0}\}$  é o espaço trivial, então  $S = \{\mathbf{0}\}$  e  $X/S = \{[\mathbf{0}]\}$ . Supondo que  $X/S$  tem co-dimensão  $k \in \mathbb{N}$ , então existe uma base  $\{W_1, \dots, W_k\}$  de  $X/S$ , onde cada  $W_j$  é uma classe de equivalência. Para cada  $j = 1, \dots, k$ , tome  $\mathbf{u}_j \in W_j$  qualquer, de modo que

$$W_j = [\mathbf{u}_j] = \mathbf{u}_j + S, \quad j = 1, \dots, k.$$

Dado  $\mathbf{u} \in X$ , temos que  $[[\mathbf{u}]] \in X/S$ , de modo que  $[[\mathbf{u}]]$  pode ser escrito como combinação linear dos elementos da base de  $X/S$ , ou seja, existem  $\lambda_1, \dots, \lambda_k \in \mathbb{K}$  tais que

$$[[\mathbf{u}]] = \lambda_1 \tilde{\odot} [[\mathbf{u}_1]] \tilde{\oplus} \dots \tilde{\oplus} \lambda_k \tilde{\odot} [[\mathbf{u}_k]] = \lambda_1 \mathbf{u}_1 + \dots + \lambda_k \mathbf{u}_k + S.$$

Dessa maneira,  $\mathbf{u}$  e  $\lambda_1 \mathbf{u}_1 + \dots + \lambda_k \mathbf{u}_k$  estão em uma mesma classe de equivalência, ou seja,

$$\mathbf{u} - \lambda_1 \mathbf{u}_1 + \dots + \lambda_k \mathbf{u}_k \in S,$$

Seja, então

$$\mathbf{v} = \mathbf{u} - \lambda_1 \mathbf{u}_1 + \dots + \lambda_k \mathbf{u}_k,$$

de maneira que  $\mathbf{v} \in S$  e

$$\mathbf{u} = \lambda_1 \mathbf{u}_1 + \dots + \lambda_k \mathbf{u}_k + \mathbf{v},$$

provando a existência da representação. A unicidade da representação de  $\mathbf{u}$  segue da unicidade da representação dos elementos de  $X/S$  em relação à base. De fato, se

$$\mathbf{u} = \lambda_1 \mathbf{u}_1 + \dots + \lambda_k \mathbf{u}_k + \mathbf{v}$$

e

$$\mathbf{u} = \lambda'_1 \mathbf{u}_1 + \dots + \lambda'_k \mathbf{u}_k + \mathbf{v}'$$

são duas representações de  $\mathbf{u}$ , com  $\lambda_1, \dots, \lambda_k, \lambda'_1, \dots, \lambda'_k \in \mathbb{K}$  e  $\mathbf{v}, \mathbf{v}' \in S$ , então o elemento  $[[\mathbf{u}]] \in X/S$  pode ser escrito como

$$[[\mathbf{u}]] = [[\lambda_1 \mathbf{u}_1 + \dots + \lambda_k \mathbf{u}_k]] = [[\lambda'_1 \mathbf{u}_1 + \dots + \lambda'_k \mathbf{u}_k]].$$

Portanto,

$$\lambda_1 \tilde{\odot} [[\mathbf{u}_1]] \tilde{\oplus} \dots \tilde{\oplus} \lambda_k \tilde{\odot} [[\mathbf{u}_k]] = \lambda'_1 \tilde{\odot} [[\mathbf{u}_1]] \tilde{\oplus} \dots \tilde{\oplus} \lambda'_k \tilde{\odot} [[\mathbf{u}_k]]$$

Pela unicidade dessa representação, visto que  $\{[[\mathbf{u}_1]], \dots, [[\mathbf{u}_k]]\}$  é uma base de  $X/S$ , deduzimos que

$$\lambda_j = \lambda'_j, \quad \forall j = 1, \dots, k.$$

Assim, deduzimos também que

$$\mathbf{v} = \mathbf{u} - \lambda_1 \mathbf{u}_1 + \dots + \lambda_k \mathbf{u}_k = \mathbf{u} - \lambda'_1 \mathbf{u}_1 + \dots + \lambda'_k \mathbf{u}_k = \mathbf{v}',$$

completando a demonstração da unicidade dessa representação. Suponha agora que  $S$  tenha dimensão  $m \in \mathbb{N}$  e seja  $\{\{\mathbf{w}_1, \dots, \mathbf{w}_m\}\}$  uma base de  $S$ . Como  $\mathbf{v} \in S$ , então  $\mathbf{v}$  pode ser escrito de maneira única em termos de  $\mathbf{w}_1, \dots, \mathbf{w}_m$ . Assim,  $\mathbf{u}$  pode ser escrito de maneira única em termos de  $\mathbf{w}_1, \dots, \mathbf{w}_m, \mathbf{u}_1, \dots, \mathbf{u}_k$ . Ou seja,  $\{\{\mathbf{u}_1, \dots, \mathbf{u}_k\}\}$  é uma base de um espaço complementar  $S' = \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$  de dimensão  $k$ , com  $X = S \oplus S'$ . Portanto,  $k$  é tanto a co-dimensão de  $S$  como a dimensão de  $X/S$ , mostrando que as duas dimensões coincidem.  $\square$

Como todos os espaços vetoriais de mesma dimensão finita são isomorfos entre si, obtemos que qualquer complementar de um subconjunto  $S$  de um espaço de dimensão finita é isomorfo a  $X/S$ .

**COROLÁRIO 2.2.** *Todo complementar de um subespaço vetorial  $S$  de um espaço vetorial  $X$  de dimensão finita é isomorfo ao espaço quociente  $X/S$ .*

**OBSERVAÇÃO 2.8.** A identidade (2.1) pode ser usada para dar uma outra demonstração à **Proposição 2.4**. A ideia é considerar  $S = U \cap V$ , mostrar que  $(U/S) \cap (V/S)$  é o subespaço trivial de  $(U + V)/S$ , que  $(U + V)/S = U/S + V/S$ , para deduzir que  $(U + V)/S = (U/S) \oplus (V/S)$  é soma direta e, assim, obter que

$$\begin{aligned} \dim(U + V) &= \dim((U + V)/S) + \dim(S) \\ &= \dim(U/S) + \dim(V/S) + \dim(S) \\ &= \dim(U) - \dim(S) + \dim(V) - \dim(S) + \dim(S) \\ &= \dim(U) + \dim(V) - \dim(S). \end{aligned}$$

### 3. Base de Hamel

Vários desses conceitos de espaços vetoriais de dimensão finita, como  $\mathbb{R}^d$  e  $\mathbb{C}^d$ , podem ser estendido para espaços vetoriais quaisquer, de dimensão infinita. Na verdade, várias definições importantes envolvem séries, que, por sua vez, requerem alguma noção de limite e, com isso, alguma topologia. Sem topologia, a definição de base que podemos trabalhar se baseia, apenas, em combinações lineares finitas. Nesse sentido, distinguimos esse conceito mais geral como base de Hamel. Vamos estender alguns dos conceitos acima para esse contexto mais geral.

**3.1. Combinação linear de elementos de um conjunto qualquer.** Primeiramente, vemos o conceito de combinação linear de vetores em um conjunto qualquer, não apenas de um conjunto finito de vetores.

**DEFINIÇÃO 3.1** (Combinação linear). *Seja  $X$  um espaço vetorial sobre um corpo  $\mathbb{K}$ . Dizemos que um elemento  $\mathbf{u} \in X$  é uma **combinação linear** de elementos de um subconjunto  $A \subset X$  quando  $\mathbf{u}$  é uma combinação linear de um conjunto finito de elementos de  $A$ , ou seja, quando existem  $\mathbf{v}_1, \dots, \mathbf{v}_n \in A$ ,  $n \in \mathbb{N}$ , e  $x_1, \dots, x_n \in \mathbb{K}$ , tais que*

$$\mathbf{u} = x_1 \mathbf{v}_1 + \dots + x_n \mathbf{v}_n.$$

**EXEMPLO 3.1** (Combinação de monômios). Um polinômio qualquer em  $\mathbb{R}$  é combinação linear de  $A = \{x^n; n = 0, 1, 2, \dots\} = \{1, x, x^2, x^3, \dots\}$ .

Também podemos estender a definição de conjuntos linearmente dependentes ou independentes a um conjunto qualquer, da seguinte forma.

### 3.2. Conjuntos linearmente dependentes e independentes.

DEFINIÇÃO 3.2 (Conjuntos LI e LD). *Seja  $X$  um espaço vetorial sobre um corpo  $\mathbb{K}$  e  $A \subset X$  um subconjunto. Dizemos que  $A$  é **linearmente independente**, ou simplesmente **LI**, quando  $A$  é não vazio e todo subconjunto finito de  $A$  é linearmente independente. Caso contrário, o conjunto é dito **linearmente dependente**, ou **LD**.*

EXEMPLO 3.2 (Monômios). No espaço de polinômios, o conjunto enumerável

$$\{1, x, x^2, x^3, \dots\}$$

é linearmente independente.

EXEMPLO 3.3 (Senoidais). No espaço vetorial  $X = \mathcal{C}(\mathbb{R})$  das funções contínuas de  $\mathbb{R}$  em  $\mathbb{R}$ , o subconjunto

$$S = \{\text{sen}(\omega t); \omega \in \mathbb{R}\}$$

é linearmente independente. Verifique!

**3.3. Base de Hamel.** Com base nisso, temos o conceito de base de Hamel:

DEFINIÇÃO 3.3 (Base de Hamel). *Seja  $X$  um espaço vetorial sobre um corpo  $\mathbb{K}$ . Dizemos que um subconjunto  $A \subset X$  é uma **base de Hamel** para  $X$  se  $A$  é linearmente independente e gera  $X$ .*

O conceito de base que estamos acostumados em  $\mathbb{R}^d$ , por exemplo, é um exemplo de base de Hamel. Naturalmente, a base de Hamel não é única, assim como a base no sentido clássico.

Os vetores de uma base de Hamel são necessariamente não nulos, portanto um espaço vetorial trivial não possui uma base de Hamel. Mas todo espaço vetorial não trivial possui uma base de Hamel, como mostra o resultado a seguir.

TEOREMA 3.1. *Todo espaço vetorial não trivial possui base de Hamel.*

A demonstração é usualmente feita através do Lema de Zorn, que é equivalente ao Axioma da Escolha. Isso tudo é visto em cursos de Análise Funcional. Veja, por exemplo, [15].

EXEMPLO 3.4. O espaço  $\mathcal{P}(\mathbb{R})$  de todos os polinômios em  $\mathbb{R}$  é um espaço vetorial de dimensão infinita enumerável, cuja base de Hamel pode ser, por exemplo, o conjunto  $\{1, x, x^2, \dots\}$ . Já o espaço de todas as funções contínuas tem base de Hamel não enumerável. De fato, todo espaço vetorial de dimensão infinita que possui

uma norma que o torna um espaço completo possui dimensão não-enumerável. Isso também é visto em cursos de Análise Funcional.

EXEMPLO 3.5. O espaço  $c_{00}$  de seqüências reais que se anulam eventualmente tem o conjunto  $\{\{e_1, e_2, \dots\}\}$  como base Hamel, onde  $e_j = (\delta_{jk})_{k \in \mathbb{N}}$ , onde  $\delta_{ij}$  é o Delta de Kronecker.

## 4. Aplicações

**4.1. Determinantes e n-volumes.** O conceito de linearidade pode ser estendido ao de multi-linearidade e nos dar uma noção de volume em dimensões arbitrárias. Esse é um dos papéis principais do conceito de determinante introduzido nos cursos iniciais de Álgebra Linear. Nesse sentido, o determinante pode ser visto como uma função em um espaço produto,

$$\det : (\mathbb{K}^n)^n = \overbrace{\mathbb{K}^n \times \dots \times \mathbb{K}^n}^{n \text{ vezes}} \rightarrow \mathbb{K}.$$

com as propriedades de ser *multilinear*, *alternante* e normalizado de tal forma a ser *unitário na base canônica*, i.e.

- (i)  $\det(\dots, \mathbf{v} + \lambda \mathbf{w}, \dots) = \det(\dots, \mathbf{v}, \dots) + \lambda \det(\dots, \mathbf{w}, \dots)$ ;
- (ii)  $\det(\dots, \mathbf{v}, \dots, \mathbf{w}, \dots) = -\det(\dots, \mathbf{w}, \dots, \mathbf{v}, \dots)$ ; e
- (iii)  $\det(\mathbf{e}_1, \dots, \mathbf{e}_n) = 1$ .

Pensando em dimensões 1, 2 e 3, a primeira e a última propriedades são características fundamentais das noções de comprimento, área e volume, respectivamente. A segunda propriedade nos dá uma noção de orientação, como se tivéssemos mostrando a frente ou o verso de uma região demarcada e orientada pelos vetores dados.

Mais precisamente, em dimensão um,  $\det(\mathbf{u})$  é simplesmente  $\det(\mathbf{u}) = x$ , para  $\mathbf{u} = (x) \in \mathbb{R}^1$ , que é o comprimento com sinal de  $x \in \mathbb{R}$ . Em dimensão dois,  $\det(\mathbf{u}, \mathbf{v})$  indica a área com sinal do paralelogramo gerado pelos vetores  $\mathbf{u} = (a, b)$  e  $\mathbf{v} = (c, d)$ , que é o paralelogramo de vértices  $(0, 0)$ ,  $(a, b)$ ,  $(a+c, b+d)$  e  $(c, d)$  e onde o sinal indica se a rotação de menor ângulo necessária para girar o vetor  $(a, b)$  na direção do vetor  $(c, d)$  é positiva (no sentido trigonométrico) ou negativa (no sentido horário). Em dimensão três,  $\det(\mathbf{u}, \mathbf{v}, \mathbf{w})$  indica a o volume, com sinal, do paralelepípedo gerado pelos vetores  $\mathbf{u}, \mathbf{v}, \mathbf{w}$ .

Com isso em mente, vamos mostrar que existe uma, e somente uma, função com as propriedades elencadas acima. Dessa forma, o determinante coincide com as noções de comprimento, área e volume *com sinal* em dimensões baixas e estende esses conceitos para dimensões mais altas. A unicidade segue do fato dela necessariamente ser uma certa função bem definida em termos das coordenadas dos vetores.

Antes de vermos isso, observe que, como consequência da propriedade (ii) acima, se dois vetores forem iguais, o determinante é nulo. De fato, trocando os dois vetores de lugar,

$$\det(\dots, \mathbf{u}, \dots, \mathbf{u}, \dots) = -\det(\dots, \mathbf{u}, \dots, \mathbf{u}, \dots),$$

mas como os vetores são iguais, então, necessariamente,

$$\det(\dots, \mathbf{u}, \dots, \mathbf{u}, \dots) = 0.$$

Isso é natural se pensarmos que o espaço gerado por esses vetores vai ter uma dimensão a menos e, com isso, ter o seu volume  $n$ -dimensional nulo.

No caso em que pelo menos um dos vetores seja nulo, então o determinante também é nulo, visto que, pela propriedade (i),

$$\det(\dots, \mathbf{0}, \dots) = \det(\dots, \lambda \mathbf{0}, \dots) = \lambda \det(\dots, \mathbf{0}, \dots),$$

para  $\lambda$  arbitrário, de modo que

$$\det(\dots, \mathbf{0}, \dots) = 0.$$

Além disso, por conta da multilinearidade, o determinante é uma função homogênea de grau  $n$ , i.e.

$$\det(\lambda \mathbf{u}_1, \dots, \lambda \mathbf{u}_n) = \lambda^n \det(\mathbf{u}_1, \dots, \mathbf{u}_n), \quad \forall \lambda \in \mathbb{R}.$$

Resumimos esses resultados da seguinte forma.

**TEOREMA 4.1.** *Dado  $n \in \mathbb{N}$  e um corpo  $\mathbb{K}$ , seja  $\det : (\mathbb{K}^n)^n \rightarrow \mathbb{K}$  uma função satisfazendo as propriedades (i), (ii) e (iii). Então, dados  $\mathbf{u}_1, \dots, \mathbf{u}_n, \mathbf{u} \in \mathbb{K}^n$  e  $\lambda \in \mathbb{K}$  arbitrários, vale*

$$\det(\mathbf{u}_1, \dots, \mathbf{u}_{i-1}, \mathbf{u}, \mathbf{u}_{i+1}, \dots, \mathbf{u}_{j-1}, \mathbf{u}, \mathbf{u}_{j+1}, \dots, \mathbf{u}_n) = 0, \quad (4.1)$$

$$\det(\mathbf{u}_1, \dots, \mathbf{u}_{j-1}, \mathbf{0}, \mathbf{u}_{j+1}, \dots, \mathbf{u}_n) = 0, \quad (4.2)$$

$$\det(\lambda \mathbf{u}_1, \dots, \lambda \mathbf{u}_n) = \lambda^n \det(\mathbf{u}_1, \dots, \mathbf{u}_n). \quad (4.3)$$

Agora estamos aptos a provar o resultado de existência e unicidade da função determinante.

**TEOREMA 4.2** (Existência e unicidade do determinante e fórmula de Leibniz). *Dado  $n \in \mathbb{N}$  e um corpo  $\mathbb{K}$ , existe uma, e somente uma, função  $\det : (\mathbb{K}^n)^n \rightarrow \mathbb{K}$  satisfazendo as propriedades (i), (ii) e (iii), que podemos escrever mais precisamente*

como

$$\begin{aligned} \det(\mathbf{u}_1, \dots, \mathbf{u}_{j-1}, \mathbf{u} + \lambda \mathbf{v}, \mathbf{u}_{j+1}, \dots, \mathbf{u}_n) \\ = \det(\mathbf{u}_1, \dots, \mathbf{u}_{j-1}, \mathbf{u}, \mathbf{u}_{j+1}, \dots, \mathbf{u}_n) \end{aligned} \quad (4.4)$$

$$+ \lambda \det(\mathbf{u}_1, \dots, \mathbf{u}_{j-1}, \mathbf{v}, \mathbf{u}_{j+1}, \dots, \mathbf{u}_n); \quad (4.5)$$

$$\begin{aligned} \det(\mathbf{u}_1, \dots, \mathbf{u}_{i-1}, \mathbf{u}_i, \mathbf{u}_{i+1}, \dots, \mathbf{u}_{j-1}, \mathbf{u}_j, \mathbf{u}_{j+1}, \dots, \mathbf{u}_n) \\ = -\det(\mathbf{u}_1, \dots, \mathbf{u}_{i-1}, \mathbf{u}_j, \mathbf{u}_{i+1}, \dots, \mathbf{u}_{j-1}, \mathbf{u}_i, \mathbf{u}_{j+1}, \dots, \mathbf{u}_n); \end{aligned} \quad (4.6)$$

$$\det(\mathbf{e}_1, \dots, \mathbf{e}_n) = 1, \quad (4.7)$$

para quaisquer  $\mathbf{u}_1, \dots, \mathbf{u}_n, \mathbf{v}, \mathbf{w} \in \mathbb{K}^n$ ,  $\lambda \in \mathbb{K}$ ,  $j \in \mathbb{N}$  com  $1 \leq j \leq n$  e  $i \in \mathbb{N}$  com  $1 \leq i < j \leq n$ . Essa função pode ser expressa através da **fórmula de Leibniz**, dada por

$$\det(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n) = \sum_{\sigma \in S_n} (-1)^{N(\sigma)} a_{\sigma(1),1} a_{\sigma(2),2} \cdots a_{\sigma(n),n}, \quad (4.8)$$

onde  $\mathbf{u}_j = (a_{1j}, \dots, a_{nj}) \in \mathbb{K}^n$ ,  $j = 1, \dots, n$  são vetores em  $\mathbb{K}^n$ ;  $S_n$  indica o conjunto de todas as  $n!$  permutações da sequência  $(1, \dots, n)$ , lembrando que uma permutação de  $(1, \dots, n)$  é uma bijeção  $\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ ; o coeficiente  $(-1)^{N(\sigma)}$  é a paridade da permutação; e o inteiro  $N(\sigma)$  é o número de inversões da permutação, i.e. o número de pares  $i, j$  em  $(1, \dots, n)$  com  $i < j$  e  $\sigma(i) > \sigma(j)$ .

**DEMONSTRAÇÃO.** Considere  $n$  vetores em  $\mathbb{K}^n$  que escrevemos como

$$\mathbf{u}_j = (a_{1j}, \dots, a_{nj}), \quad j = 1, \dots, n.$$

Em termos dos elementos da base canônica, temos

$$\mathbf{u}_j = a_{1j}\mathbf{e}_1 + \cdots + a_{nj}\mathbf{e}_n.$$

Usando a multilinearidade de  $\det$ , temos

$$\begin{aligned} \det(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n) &= \sum_{i_1=1}^n a_{i_1,1} \det(\mathbf{e}_{i_1}, \mathbf{u}_2, \dots, \mathbf{u}_n) \\ &= \sum_{i_1=1}^n \sum_{i_2=1}^n a_{i_1,1} a_{i_2,2} \det(\mathbf{e}_{i_1}, \mathbf{e}_{i_2}, \dots, \mathbf{u}_n) \\ &= \cdots \\ &= \sum_{i_1=1}^n \sum_{i_2=1}^n \cdots \sum_{i_n=1}^n a_{i_1,1} a_{i_2,2} \cdots a_{i_n,n} \det(\mathbf{e}_{i_1}, \mathbf{e}_{i_2}, \dots, \mathbf{e}_{i_n}). \end{aligned}$$

Como o determinante se anula quando dois de seus argumentos são vetores iguais, então sobram apenas as combinações com índices  $i_1, \dots, i_n$  distintos, ou seja, apenas as permutações de  $1, \dots, n$ . Assim, escrevemos

$$\det(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n) = \sum_{\sigma \in S_n} a_{\sigma(1),1} a_{\sigma(2),2} \cdots a_{\sigma(n),n} \det(\mathbf{e}_{\sigma(1)}, \mathbf{e}_{\sigma(2)}, \dots, \mathbf{e}_{\sigma(n)}),$$

onde  $S_n$  é o conjunto de todas essas permutações.

Cada determinante  $\det(\mathbf{e}_{\sigma(1)}, \mathbf{e}_{\sigma(2)}, \dots, \mathbf{e}_{\sigma(n)})$  é igual a 1 ou  $-1$ , ou, mais precisamente, é igual à paridade  $\text{sgn}(\sigma) = (-1)^{N(\sigma)}$  da permutação, que depende do número de inversões da permutação, que é o número de elementos  $i < j$  em  $(1, \dots, n)$  tais que  $\sigma(i) > \sigma(j)$ . O número de inversões nos dá o número de vezes que temos que trocar os vetores para reordená-los na ordem  $\mathbf{e}_1, \dots, \mathbf{e}_n$ . Nesta ordem, o determinante é igual a 1. A cada inversão, multiplicamos o resultado por  $-1$ , graças à propriedade do determinante ser alternante. Assim,

$$\det(\mathbf{e}_{\sigma(1)}, \mathbf{e}_{\sigma(2)}, \dots, \mathbf{e}_{\sigma(n)}) = (-1)^{N(\sigma)}.$$

Dessa maneira, obtemos a fórmula de Leibniz (4.8).  $\square$

O determinante pode ser usado para determinar se um conjunto de  $n$  vetores é uma base para o espaço  $\mathbb{K}^n$ .

**TEOREMA 4.3.** *Sejam  $\mathbf{u}_1, \dots, \mathbf{u}_n \in \mathbb{K}^n$ ,  $n \in \mathbb{N}$ . Então esses vetores formam uma base para  $\mathbb{K}^n$  se, e somente se,  $\det(\mathbf{u}_1, \dots, \mathbf{u}_n) \neq 0$ .*

**DEMONSTRAÇÃO.** Como são  $n$  vetores e o espaço  $\mathbb{K}^n$  é de dimensão  $n$  então eles formam uma base se, e somente se, são linearmente independentes. Se os vetores não formam uma base, então é porque eles são linearmente dependentes. Nesse caso, podemos escrever um dos vetores como combinação linear dos outros. Sem perda de generalidade, vamos assumir que

$$\mathbf{u}_n = a_1 \mathbf{u}_1 + \cdots + a_{n-1} \mathbf{u}_{n-1}.$$

Nesse caso, usando a propriedade de multilinearidade,

$$\begin{aligned} \det(\mathbf{u}_1, \dots, \mathbf{u}_n) &= \det(\mathbf{u}_1, \dots, \mathbf{u}_{n-1}, a_1 \mathbf{u}_1 + \cdots + a_{n-1} \mathbf{u}_{n-1}) \\ &= a_1 \det(\mathbf{u}_1, \dots, \mathbf{u}_{n-1}, \mathbf{u}_1) + \cdots + a_{n-1} \det(\mathbf{u}_1, \dots, \mathbf{u}_{n-1}, \mathbf{u}_{n-1}). \end{aligned}$$

Como o determinante se anula quando dois vetores são iguais e como todos os termos do lado direito acima envolve determinantes com dois vetores iguais, então

$$\det(\mathbf{u}_1, \dots, \mathbf{u}_n) = 0.$$

Ou seja, se os vetores são linearmente dependentes, então o determinante é nulo. Dito de outra forma, se o determinante não for nulo, então os vetores têm que ser linearmente independentes e, portanto, formar uma base.

Para a recíproca, vamos assumir que eles formam uma base. Então eles geram o espaço todo. Em particular, todos os vetores da base canônica podem ser expressos em termos dos vetores  $\mathbf{u}_1, \dots, \mathbf{u}_n$ . Escrevamos

$$\mathbf{e}_i = b_{i1}\mathbf{u}_1 + \dots + b_{in}\mathbf{u}_n,$$

para  $i = 1, \dots, n$  e coeficientes apropriados  $b_{ij}$ ,  $i, j = 1, \dots, n$ . Usando a multilinearidade e a propriedade do determinante ser alternante, obtemos o análogo da fórmula de Leibniz (4.8), mas em termos dos determinantes de permutações dos vetores  $\mathbf{u}_1, \dots, \mathbf{u}_n$ , i.e.

$$\det(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n) = \sum_{\sigma \in S_n} b_{\sigma(1),1} b_{\sigma(2),2} \dots b_{\sigma(n),n} \det(\mathbf{u}_{\sigma(1)}, \dots, \mathbf{u}_{\sigma(n)}).$$

Caso  $\det(\mathbf{u}_1, \dots, \mathbf{u}_n) = 0$ , então  $\det(\mathbf{u}_{\sigma(1)}, \dots, \mathbf{u}_{\sigma(n)}) = 0$  para qualquer permutação em  $S_n$ , de modo que  $\det(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n) = 0$ , o que é uma contradição. Logo, necessariamente  $\det(\mathbf{u}_1, \dots, \mathbf{u}_n) \neq 0$ , completando a demonstração.  $\square$

**OBSERVAÇÃO 4.1.** Em um espaço vetorial qualquer  $X$  de dimensão finita sobre um corpo  $\mathbb{K}$ , podemos construir uma função  $d : X^n = \overbrace{X \times \dots \times X}^{n \text{ vezes}} \rightarrow \mathbb{K}$  com as propriedades de multilinearidade e alternância, mas a unicidade depende da escolha de uma base  $\mathcal{b} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  de  $X$  cujo  $n$ -volume definimos como sendo  $d(\mathbf{w}_1, \dots, \mathbf{w}_n) = 1$  (ou, mais geralmente, qualquer escalar não nulo). Nesse caso, obtemos que  $d(\mathbf{u}, \dots, \mathbf{u}_n) = \det([\mathbf{u}_1]_{\mathcal{b}}, \dots, [\mathbf{u}_n]_{\mathcal{b}})$ . Mas isso depende fortemente da escolha da base. Por esse motivo, não costumamos definir o determinante em um espaço vetorial arbitrário. Veremos, no entanto, na [Seção 3](#), que isso não nos impede de definir o determinante de um operador linear em um espaço vetorial arbitrário de dimensão finita, porque, no fundo, o determinante de um operador mede um  $n$ -volume *relativo*, ou seja, mede a taxa de mudança de  $n$ -volume, eliminando a dependência na escolha da base. Isso ficará mais claro no devido momento.

**4.2. Representação computacional de matrizes especiais.** Na [Seção 2.5](#), vimos algumas matrizes quadradas especiais, como matrizes simétricas e matrizes diagonais. Como dito, essas formas especiais podem ser exploradas em pacotes computacionais para economizar espaço de memória e para acelerar certos cálculos numéricos, tanto as operações essenciais de multiplicação por escalar e adição, como multiplicação (composição) de matrizes e diversos tipos de fatoração.

Abaixo, por exemplo, uma lista de alguns tipos especiais de matrizes, definidos no pacote padrão `LinearAlgebra.jl` da linguagem de programação Julia:

```
Bidiagonal{T,V<:AbstractVector{T}} <: AbstractMatrix{T}
Diagonal{T,V<:AbstractVector{T}} <: AbstractMatrix{T}
Tridiagonal{T,V<:AbstractVector{T}} <: AbstractMatrix{T}
Hermitian{T,S<:AbstractMatrix{<:T}} <: AbstractMatrix{T}
Symmetric{T,S<:AbstractMatrix{<:T}} <: AbstractMatrix{T}
SymTridiagonal{T, V<:AbstractVector{T}} <: AbstractMatrix{T}
UpperHessenberg{T,S<:AbstractMatrix{T}} <: AbstractMatrix{T}
```

Por exemplo, a matriz diagonal guarda apenas um vetor com os elementos da diagonal, o que pode ser visto através da sua estrutura:

```
struct Diagonal{T,V<:AbstractVector{T}} <: AbstractMatrix{T}
    diag::V

    function Diagonal{T,V}(diag) where {T,V<:AbstractVector{T}}
        require_one_based_indexing(diag)
        new{T,V}(diag)
    end
end
```

Por sua vez, a matriz tridiagonal simétrica guarda, na memória, apenas dois vetores, um para diagonal e outra para a superdiagonal, que serve também para a subdiagonal. Isso pode ser visto através da sua estrutura (omitindo o construtor interno, para melhor visualização):

```
struct SymTridiagonal{T, V<:AbstractVector{T}} <: AbstractMatrix{T}
    dv::V # diagonal
    ev::V # superdiagonal
end
```

Podemos criar uma matriz tridiagonal simétrica fornecendo esses dois vetores:

```
julia> SymTridiagonal([1, 2, 3, 4], [5, 6, 7])
4×4 SymTridiagonal{Int64, Vector{Int64}}:
 1  5  .  .
 5  2  6  .
 .  6  3  7
 .  .  7  4
```

Algumas estruturas guardam um espaço a mais para a alocação temporária de variáveis necessárias para certas operações, ganhando eficiência nos cálculos, como a estrutura para matriz tridiagonal, que guarda um vetor para pivoteamento na decomposição LU (que veremos posteriormente):

```

struct Tridiagonal{T,V<:AbstractVector{T}} <: AbstractMatrix{T}
    dl::V      # sub-diagonal
    d::V       # diagonal
    du::V      # sup-diagonal
    du2::V     # supsup-diagonal for pivoting in LU
end

```

Abaixo alguns outros exemplos de construção de matrizes especiais.

```

julia> Diagonal([1 2; 3 4])
2×2 Diagonal{Int64, Vector{Int64}}:
 1  .
 .  4

julia> Diagonal([1, 2, 3])
3×3 Diagonal{Int64, Vector{Int64}}:
 1  .  .
 .  2  .
 .  .  3

julia> Bidiagonal(ones(Int, 4, 4), :U)
4×4 Bidiagonal{Int64, Vector{Int64}}:
 1  1  .  .
 .  1  1  .
 .  .  1  1
 .  .  .  1

julia> Tridiagonal([21, 32, 43], [11, 22, 33, 44], [12, 23, 34])
4×4 Tridiagonal{Int64, Vector{Int64}}:
 11  12  .  .
 21  22  23  .
 .  32  33  34
 .  .  43  44

julia> Symmetric([1 2; 3 4])
2×2 Symmetric{Int64, Matrix{Int64}}:
 1  2
 2  4

julia> Symmetric([1 + 1im 1 + 2im; 1 + 3im 1 + 4im])
2×2 Symmetric{Complex{Int64}, Matrix{Complex{Int64}}}:
 1+1im  1+2im

```

```

1+2im 1+4im

julia> Hermitian([1 + 1im 1 + 2im; 1 + 3im 1 + 4im])
2×2 Hermitian{Complex{Int64}, Matrix{Complex{Int64}}}:
 1+0im 1+2im
 1-2im 1+0im

julia> UpperHessenberg(reshape(1:25, 5, 5))
5×5 UpperHessenberg{Int64, Base.ReshapedArray{Int64, 2, UnitRange{Int64}, Tuple{
}}}:
 1  6  11  16  21
 2  7  12  17  22
 .  8  13  18  23
 .  .  14  19  24
 .  .  .  20  25

```

Conforme mencionado, essas estruturas podem ser exploradas também para acelerar cálculos numéricos. Por exemplo, a seguinte linha define como é a adição entre duas matrizes diagonais. Observe que ela simplesmente adiciona os vetores que definem a diagonal de cada matriz (armazenada no campo `diag`) e cria uma nova matriz diagonal com o vetor resultante:

```
(+) (Da::Diagonal, Db::Diagonal) = Diagonal(Da.diag + Db.diag)
```

Isso explicita o fato do conjunto das matrizes diagonais ser um subespaço linear, visto que a adição de duas matrizes diagonais gera uma nova matriz diagonal. Idem para a multiplicação por escalar:

```
(*) (x::Number, D::Diagonal) = Diagonal(x * D.diag)
```

Vejamos um exemplo:

```

julia> Da = Diagonal([1, 2, 3])
3×3 Diagonal{Int64, Vector{Int64}}:
 1  .  .
 .  2  .
 .  .  3

julia> Db = Diagonal([4, 5, 6])
3×3 Diagonal{Int64, Vector{Int64}}:
 4  .  .
 .  5  .
 .  .  6

```

```

julia> Da + Db
3×3 Diagonal{Int64, Vector{Int64}}:
 5  .  .
 .  7  .
 .  .  9

julia> λ = 2
2

julia> λ * Da
3×3 Diagonal{Int64, Vector{Int64}}:
 2  .  .
 .  4  .
 .  .  6

```

O código onde essa adição especial é definida pode ser encontrado através das macros `@which` ou `@edit`

```

julia> @which Da + Db
+(Da::Diagonal, Db::Diagonal)
 @ LinearAlgebra /Applications/Julia-1.9.app/Contents/Resources/julia/share/
  julia/stdlib/v1.9/LinearAlgebra/src/diagonal.jl:184

```

Junto com essa definição, encontramos várias outras operações otimizadas envolvendo matrizes diagonais:

```

(==) (Da::Diagonal, Db::Diagonal) = Da.diag == Db.diag
(-) (A::Diagonal) = Diagonal(-A.diag)
(+) (Da::Diagonal, Db::Diagonal) = Diagonal(Da.diag + Db.diag)
(-) (Da::Diagonal, Db::Diagonal) = Diagonal(Da.diag - Db.diag)

for f in (:+, :-)
    @eval function $f(D::Diagonal, S::Symmetric)
        return Symmetric($f(D, S.data), sym_uplo(S.uplo))
    end
    @eval function $f(S::Symmetric, D::Diagonal)
        return Symmetric($f(S.data, D), sym_uplo(S.uplo))
    end
    @eval function $f(D::Diagonal{<:Real}, H::Hermitian)
        return Hermitian($f(D, H.data), sym_uplo(H.uplo))
    end
    @eval function $f(H::Hermitian, D::Diagonal{<:Real})
        return Hermitian($f(H.data, D), sym_uplo(H.uplo))
    end
end
end

```

```

(*) (x::Number, D::Diagonal) = Diagonal(x * D.diag)
(*) (D::Diagonal, x::Number) = Diagonal(D.diag * x)
(/) (D::Diagonal, x::Number) = Diagonal(D.diag / x)
(\) (x::Number, D::Diagonal) = Diagonal(x \ D.diag)
(^) (D::Diagonal, a::Number) = Diagonal(D.diag .^ a)
(^) (D::Diagonal, a::Real) = Diagonal(D.diag .^ a) # for disambiguation
(^) (D::Diagonal, a::Integer) = Diagonal(D.diag .^ a) # for disambiguation
Base.literal_pow(::typeof(^), D::Diagonal, valp::Val) =
    Diagonal(Base.literal_pow(^, D.diag, valp)) # for speed
Base.literal_pow(::typeof(^), D::Diagonal, ::Val{-1}) = inv(D) # for
    disambiguation

```

Para efeito de comparação, o teste a seguir exibe o tempo de processamento e a alocação de memória para a adição de duas matrizes diagonais  $1000 \times 1000$  guardadas de maneira completa, como se fossem matrizes densas, preenchida com zeros fora da diagonal, e o tempo e a alocação obtidos explorando a estrutura diagonal. Observe a imensa diferença entre os dois tratamentos. Da ordem de 1000 vezes mais rápido e menos memória.

```

julia> using BenchmarkTools

julia> Da = Diagonal(rand(Int, 1000)); Db = Diagonal(rand(Int, 1000));

julia> A = Matrix(Da); B = Matrix(Db); # matrizes completas

julia> typeof(A)
Matrix{Int64} (alias for Array{Int64, 2})

julia> typeof(Da)
Diagonal{Int64, Vector{Int64}}

julia> @btime (+) ($A, $B);
304.042 μs (2 allocations: 7.63 MiB)

julia> @btime (+) ($Da, $Db);
231.436 ns (1 allocation: 7.94 KiB)

```

Poucas linguagens de programação exploram tão bem essas estruturas. Julia é um caso particular. Além das estruturas acima, diversas outras matrizes especiais são definidas e otimizadas em outros pacotes.

## 5. Exercícios

### Exercícios

**5.1.** Considere o conjunto

$$X = \{u = (r, \cos(r)) \in \mathbb{R}^2; r \in \mathbb{R}\}$$

munido das operações de adição e multiplicação por escalar em  $\mathbb{R}$  definidas por

$$\begin{aligned}(r_1, \cos(r_1)) + (r_2, \cos(r_2)) &= (r_1 + r_2, \cos(r_1 + r_2)), \\ \lambda(r, \cos(r)) &= (\lambda r, \cos(\lambda r)).\end{aligned}$$

Verifique se  $X$  é um espaço vetorial sobre  $\mathbb{R}$  ou não. Justifique.

**5.2.** Sejam  $\mathbf{u}, \mathbf{v}, \mathbf{w}$  vetores em um espaço vetorial. Demonstre, passo a passo, que se  $\mathbf{u} + \mathbf{w} = \mathbf{v} + \mathbf{w}$ , então  $\mathbf{u} = \mathbf{v}$ , indicando qual axioma de espaço vetorial está sendo usado em cada passo.

**5.3.** Seja  $n \in \mathbb{N}$  e  $\mathbf{v}$  um vetor em um espaço vetorial sobre o corpo  $\mathbb{R}$ . Demonstre que  $n\mathbf{v} = \mathbf{v} + \cdots + \mathbf{v}$ ,  $n$  vezes, indicando, passo a passo, qual axioma de espaço vetorial está sendo usado.

**5.4.** Considere o conjunto  $\mathbb{R}^2$  munido da multiplicação por escalar  $\lambda(x, y) = (\lambda x, \lambda y)$  usual, mas com a adição definida por

$$(x_1, y_1) \oplus (x_2, y_2) = (x_1 + y_2, x_2 + y_1).$$

Verifique quais axiomas de espaço vetorial continuam válidos e quais não são.

**5.5.** Considere o conjunto  $\mathbb{R}^2$  munido da multiplicação por escalar  $\lambda(x, y) = (\lambda x, \lambda y)$  usual, mas com a adição definida pelo valor médio

$$(x_1, y_1) \oplus (x_2, y_2) = \left( \frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2} \right).$$

Verifique quais axiomas de espaço vetorial continuam válidos e quais não são.

**5.6.** Verifique que o espaço complexificado definido no [Exemplo 1.8](#) é, de fato, um espaço vetorial complexo e identifique o elemento neutro da adição.

**5.7.** Seja  $X$  um espaço vetorial sobre os reais. Prove que

- (a) Todo subespaço afim de  $X$  é convexo.
- (b) Se um subespaço afim de  $X$  é um cone, então é um subespaço.
- (c) Todo subconjunto convexo de  $X$  é um conjunto de formato estrela.
- (d) Se  $C$  é um subconjunto convexo de  $X$ ,  $\mathbf{u}_1, \dots, \mathbf{u}_n \in C$  e  $0 \leq \theta_1, \dots, \theta_n \leq 1$  com  $\sum_j \theta_j = 1$ , para algum  $n \in \mathbb{N}$ , então  $\sum_{j=1}^n \theta_j \mathbf{u}_j \in C$ .
- (e) Um subconjunto  $C$  é um cone convexo não vazio simétrico (i.e.  $\mathbf{u} \in C \Rightarrow -\mathbf{u} \in C$ ) se, e somente se, é um subespaço.

**5.8.** Considerando os objetos abaixo como subconjuntos de um espaço vetorial sobre os reais, prove que

- (a) A união arbitrária de subespaços vetoriais pode não ser um subespaço vetorial;
- (b) A interseção arbitrária de subespaços vetoriais é um subespaço vetorial;
- (c) A união arbitrária de cones é um cone;
- (d) A interseção arbitrária de cones é um cone;
- (e) A união arbitrária de convexos pode não ser convexa;
- (f) A interseção arbitrária de convexos é um convexo;
- (g) A união arbitrária de conjuntos de formato estrela pode não ter formato estrela;
- (h) A interseção arbitrária de conjuntos de formato estrela pode não ser de formato estrela;

**5.9.** Prove que o conjunto  $C = \{f \in \mathcal{C}(I, \mathbb{R}); \int_I f(x) dx \geq 0\}$  é um cone convexo no espaço vetorial  $X = \mathcal{C}(I, \mathbb{R})$  sobre  $\mathbb{R}$ .

**5.10.** Prove que o conjunto de funções sub-harmônicas

$$C = \{f \in \mathcal{C}^2(\Omega, \mathbb{R}); \Delta f \geq 0\}$$

em um domínio  $\Omega \subset \mathbb{R}^n$ , com  $n \in \mathbb{N}$ , é um cone convexo no espaço vetorial  $\mathcal{C}^2(\Omega, \mathbb{R})$  das funções duas vezes continuamente diferenciáveis em  $\Omega$ . Lembrando que  $\Delta f = \partial_{x_1}^2 f + \dots + \partial_{x_n}^2 f$  é o operador laplaciano.

**5.11.** A *envoltória convexa* de um subconjunto  $A$  em um espaço vetorial  $X$  é o menor conjunto convexo de  $X$  que contém  $A$ , ou seja, é um subconjunto convexo de  $X$  que contém  $A$  e está contido em qualquer outro subconjunto convexo de  $X$  que contém  $A$ . Prove que todo subconjunto de  $X$  possui uma envoltória convexa.

**5.12.** Seja  $X$  um espaço vetorial e seja  $C \in X$  um cone em  $X$ . Prove que  $C$  é convexo se, e somente se,  $C$  é invariante por adição, i.e. se  $\mathbf{u}, \mathbf{v} \in C$ , então  $\mathbf{u} + \mathbf{v} \in C$ . O que falta para que  $C$  seja um subespaço?

**5.13.** Lembremos que uma *relação de ordem parcial* em um conjunto é uma relação que é reflexiva ( $\mathbf{u} \leq \mathbf{u}$  para todo  $\mathbf{u}$  no conjunto), antisimétrica ( $\mathbf{u} \leq \mathbf{v}$  e  $\mathbf{v} \leq \mathbf{u}$  se, e somente se,  $\mathbf{u} = \mathbf{v}$ ) e transitiva ( $\mathbf{u} \leq \mathbf{v}$  e  $\mathbf{v} \leq \mathbf{w}$  implica em  $\mathbf{u} \leq \mathbf{w}$ ). Em um espaço vetorial real  $X$ , uma relação de ordem parcial  $\leq$  em  $X$  é dita *compatível com a estrutura vetorial* quando, para  $\mathbf{u}, \mathbf{v}, \mathbf{w} \in X$  e  $\lambda > 0$ , temos  $\mathbf{u} - \mathbf{w} \leq \mathbf{v} - \mathbf{w}$  e  $\lambda \mathbf{u} \leq \lambda \mathbf{v}$ , sempre que  $\mathbf{u} \leq \mathbf{v}$ . Nesse caso, dizemos que  $X$  munido de  $\leq$  é um *espaço vetorial com relação de ordem*. Um exemplo é o espaço  $\mathbb{R}^n$  munido da relação  $(x_1, \dots, x_n) \leq (y_1, \dots, y_n)$  se, e somente se,  $x_j \leq y_j$ . Agora, dado um conjunto  $C \subset X$  em um espaço vetorial real  $X$ , defina  $\mathbf{u} \leq \mathbf{v}$  se, e somente se  $\mathbf{v} - \mathbf{u} \in C$ . Mostre que essa relação é uma relação de ordem parcial se, e somente se,  $C$  é um cone.

- 5.14.** Identifique quais subconjuntos abaixo são subespaços vetoriais
- (a)  $S = \{(x, y, z) \in \mathbb{R}^3; x + 2y + 3z = 4\}$  em  $X = \mathbb{R}^3$ ;
  - (b)  $S = \{(x, y, z, w) \in \mathbb{R}^4; w = x^2 + y^2 + z^2 + 1\}$  em  $X = \mathbb{R}^4$ ;
  - (c)  $S = \{p \in \mathcal{P}(\mathbb{R}); p'(x) = p(x)\}$  em  $X = \mathcal{P}(\mathbb{R})$ ;
  - (d)  $S = \{p \in \mathcal{P}(\mathbb{R}); p'(x) = p(x), p(0) = 1\}$  em  $X = \mathcal{P}(\mathbb{R})$ ;
  - (e)  $S = \{f \in \mathcal{C}(I, \mathbb{R}); \int_I f(x) dx = 1\}$  em  $X = \mathcal{C}(I, \mathbb{R})$ , onde  $I \subset \mathbb{R}$  é um intervalo;
  - (f)  $S = \{A \in \mathbb{C}^{n \times n}; A + \bar{A} = I\}$  em  $X = \mathbb{C}^{n \times n}$ , onde  $\bar{A}$  é a matriz conjugada a  $A$ , ou seja, cada coeficiente de  $\bar{A}$  é o complexo conjugado do coeficiente correspondente em  $A$ , e onde  $I$  é a matriz identidade.
  - (g)  $S = \{f \in \mathcal{C}(\mathbb{R}); \exists \omega \in \mathbb{R}, f(x) = \text{sen}(\omega x) + 1; \forall x \in \mathbb{R}\}$  em  $X = \mathcal{C}(\mathbb{R})$ , ou seja, é o conjunto das funções  $f(x) = \text{sen}(\omega x) + 1$  para alguma frequência  $\omega$  arbitrária.
- 5.15.** Considere um conjunto  $\mathcal{a} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$  de vetores em um espaço vetorial  $X$ . Prove que  $\mathcal{b} = \{\mathbf{u}_2 - \mathbf{u}_1, \mathbf{u}_3 - \mathbf{u}_2, \mathbf{u}_1 - \mathbf{u}_3\}$  é LD, independentemente de  $\mathcal{a}$  ser LD ou LI.
- 5.16.** Considere um conjunto de vetores  $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$  em um espaço vetorial  $X$ . Prove que  $\text{span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} = \text{span}\{\mathbf{u}_1, \mathbf{u}_2 - \mathbf{u}_1, \mathbf{u}_3 - \mathbf{u}_1\}$ .
- 5.17.** Sejam  $\mathbf{u}_1$  e  $\mathbf{u}_2$  dois vetores em um certo espaço vetorial. Sabendo que  $\{\mathbf{u}_1 - \mathbf{u}_2, \mathbf{u}_2 - \mathbf{u}_1\}$  é LD, podemos afirmar que  $\{\mathbf{u}_1, \mathbf{u}_2\}$  é LD também ou não? Justifique.
- 5.18.** Considere os vetores complexos  $\mathbf{u} = (1, 1)$  e  $\mathbf{v} = (i, i)$ . Prove que  $\{\mathbf{u}, \mathbf{v}\}$  é um conjunto LD quando considerados no espaço vetorial  $X = \mathbb{C}^2$  sobre o corpo dos complexos  $\mathbb{K} = \mathbb{C}$  mas é LI quando considerado no espaço vetorial  $W = \mathbb{C}^2$  sobre o corpo dos reais  $\mathbb{R}$ .
- 5.19.** Demonstre que se  $\mathbf{u}_1, \dots, \mathbf{u}_k$  são vetores LI em um espaço vetorial  $X$  sobre um corpo  $\mathbb{K}$  e se  $\lambda_1, \dots, \lambda_k \in \mathbb{K}$  são escalares não nulos, então  $\lambda_1 \mathbf{u}_1, \dots, \lambda_k \mathbf{u}_k$  também são LI.
- 5.20.** Sejam  $S$  um subespaço vetorial de um espaço vetorial  $X$  e seja  $A \subset X$  um subconjunto qualquer não vazio. Suponha que todo  $\mathbf{u} \in S + A$  pode ser escrito de maneira única como soma  $\mathbf{u} = \mathbf{v} + \mathbf{a}$  de vetores  $\mathbf{v} \in S$  e  $\mathbf{a} \in A$ . Podemos deduzir que  $A$  também é um subespaço vetorial de  $X$  e que  $X = S \oplus A$ ?
- 5.21.** Seja  $W$  um subespaço de um espaço vetorial  $X$ . Demonstre que  $W = S_1 \oplus \dots \oplus S_k$  é soma direta de subespaços  $S_1, \dots, S_k$  se, e somente se,  $W = S_1 + \dots + S_k$  e  $S_i \cap \left(\sum_{j \neq i} S_j\right) = \{\mathbf{0}\}$ , para todo  $i = 1, \dots, k$ .
- 5.22.** Seja  $W$  um subespaço de um espaço vetorial  $X$ . Demonstre que  $W = S_1 \oplus \dots \oplus S_k$  para subespaços  $S_1, \dots, S_k$  se, e somente se,  $W = S_1 + \dots + S_k$  e o

vetor nulo se escreve de forma única como soma de vetores nesses espaços, i.e. se  $\mathbf{0} = \mathbf{u}_1 + \dots + \mathbf{u}_k$ , com  $\mathbf{u}_j \in S_j$ ,  $j = 1, \dots, k$ , então  $\mathbf{u}_j = \mathbf{0}$ , para todo  $j = 1, \dots, k$ .

**5.23.** Ache um subespaço complementar do subespaço

$$S = \{u \in \mathcal{C}(I, \mathbb{R}); \int_I u(x) \, dx = 0\}$$

do espaço vetorial real  $X = \mathcal{C}(I, \mathbb{R})$ , onde  $I \subset \mathbb{R}$  é um intervalo e deduza que  $S$  tem co-dimensão um.

**5.24.** Sejam  $X$  e  $Y$  dois espaços vetoriais sobre um mesmo corpo  $\mathbb{K}$  e suponha que os dois tenham dimensão finita. Demonstre que o produto cartesiano  $X \times Y = \{(v, w); v \in X, w \in Y\}$  é um espaço vetorial de dimensão finita, com

$$\dim(X \times Y) = \dim(X) + \dim(Y).$$

Observe que o mesmo vale caso  $X$  e  $Y$  sejam subespaços vetoriais de um mesmo espaço vetorial ou de espaços vetoriais diferentes sobre o mesmo corpo.

**5.25.** Complete a demonstração da fórmula  $\dim(U + W) = \dim(U) + \dim(W) - \dim(U \cap W)$  da [Proposição 2.4](#), partindo de uma base de  $U + W$  e completando até bases de  $U$  e  $W$ .

**5.26.** Prove que um subespaço  $S$  de um espaço vetorial  $X$  tem co-dimensão finita  $k \in \mathbb{N}$  se, e somente se, existem vetores  $\mathbf{v}_1, \dots, \mathbf{v}_k \in X$  tais que todo  $\mathbf{u} \in X$  pode ser representado de maneira única como uma combinação linear

$$\mathbf{u} = \mathbf{w} + \lambda_1 \mathbf{v}_1 + \dots + \lambda_k \mathbf{v}_k,$$

onde  $\lambda_1, \dots, \lambda_k \in \mathbb{K}$  e  $\mathbf{w} \in S$ .

**5.27.** Dado um subespaço  $S$  de um espaço vetorial  $X$ , verifique que a relação  $\mathbf{u} \sim \mathbf{v} \Leftrightarrow \mathbf{u} - \mathbf{v} \in S$  entre vetores  $\mathbf{u}$  e  $\mathbf{v}$  de  $X$  é, de fato, uma relação de equivalência, i.e. satisfaz as propriedades de (i) reflexividade:  $\mathbf{u} \sim \mathbf{u}$ ; (ii) simetria:  $\mathbf{u} \sim \mathbf{v}$  se, e somente se,  $\mathbf{v} \sim \mathbf{u}$ ; e (iii) transitividade:  $\mathbf{u} \sim \mathbf{v}$  e  $\mathbf{v} \sim \mathbf{w}$  implica em  $\mathbf{u} \sim \mathbf{w}$ .

**5.28.** Considere  $n + 1$  pontos distintos  $x_0, \dots, x_n$  na reta. Prove que os polinômios de Lagrange, definidos por

$$\ell_j(x) = \frac{\prod_{i \neq j} (x - x_i)}{\prod_{i \neq j} (x_j - x_i)},$$

formam uma base  $\mathcal{P}_n(\mathbb{R})$ .

- 5.29.** Seja  $p \in \mathcal{P}_3(\mathbb{R}) \setminus \mathcal{P}_2(\mathbb{R})$  um polinômio de grau três. Demonstre que  $\mathcal{b} = \{p, p', p'', p'''\}$  é uma base de  $\mathcal{P}_3(\mathbb{R})$ , onde  $p'$ ,  $p''$  e  $p'''$  são as derivadas de  $p = p(x)$  de ordem um, dois e três em relação a  $x$ .
- 5.30.** Considere o subconjunto

$$S = \{u(x) = C_0 e^x \cos(\theta + x); C_0, \theta \in \mathbb{R}\}$$

do espaço vetorial  $X = \mathbb{R}^{\mathbb{R}}$  das funções reais na reta. Demonstre que  $S$  é um subespaço de  $X$ , encontre uma base para  $S$  e determine a sua dimensão.

- 5.31.** Demonstre que  $S = \{p \in \mathcal{P}_3(\mathbb{R}); p'(1) = 0\}$  é um subespaço vetorial de  $\mathcal{P}_3(\mathbb{R})$  e que  $\mathcal{a} = \{1, (x-1)^2; (x-1)^2(x-2)\}$  é uma base para  $S$ .
- 5.32.** Determine o subespaço complementar, em  $\mathcal{P}_3(\mathbb{R})$ , do subespaço  $S = \{p \in \mathcal{P}_3(\mathbb{R}); p'(1) = 0\}$  considerado no [Exercício 5.31](#).
- 5.33.** Ache uma base para o subespaço  $S = \{p \in \mathcal{P}_4(\mathbb{R}); p''(2) = 0\}$  e encontre um subespaço complementar  $S'$  de  $S$  em  $\mathcal{P}_4(\mathbb{R})$ .
- 5.34.** Considere o espaço vetorial  $X = \mathcal{C}^1(\mathbb{R})$  das funções contínuas diferenciáveis em  $\mathbb{R}$ . Seja  $f \in X$  uma função não nula tal que  $\{f, f'\}$  seja linearmente dependente. Determine  $f$  (a menos de constantes).
- 5.35.** Seja  $X = \mathbb{R}^{2 \times 2}$  o espaço das matrizes reais quadradas. Considere o subespaço das matrizes simétricas  $S = \{A \in \mathbb{R}^{2 \times 2}; A^{\text{tr}} = A\}$ , i.e. da forma  $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ . Ache a dimensão de  $S$ . Encontre, ainda, um subespaço  $S'$  complementar a  $S$ , i.e. tal que  $X = S \oplus S'$ .
- 5.36.** Sejam  $U, V$  subespaços de um certo espaço vetorial  $X$ . Considerando  $S = U \cap V$ , prove que  $(U/S) \cap (V/S)$  é o subespaço trivial de  $X/S$  (veja [Observação 2.8](#)).
- 5.37.** Sejam  $U, V$  subespaços vetoriais de um certo espaço vetorial. Considerando  $S = U \cap V$ , prove que  $(U+V)/S = (U/S) \oplus (V/S)$  (veja [Observação 2.8](#)).
- 5.38.** Sejam  $U, V, W$  subespaços vetoriais de dimensão finita de um certo espaço vetorial. Prove ou dê um contra-exemplo para a identidade

$$\begin{aligned} \dim(U + V + W) &= \dim(U) + \dim(V) + \dim(W) \\ &\quad - \dim(U \cap V) - \dim(U \cap W) - \dim(V \cap W) \\ &\quad + \dim(U \cap V \cap W). \end{aligned}$$

- 5.39.** Sejam  $U$  e  $V$  subespaços vetoriais com interseção trivial,  $U \cap V = \{\mathbf{0}\}$ , em um espaço vetorial  $X$ . Sejam  $\mathcal{b}_U$  e  $\mathcal{b}_V$  bases de  $U$  e  $V$ . Mostre que  $\mathcal{b}_U \cup \mathcal{b}_V$  é uma base de  $U \oplus V$ .

**5.40.** Seja  $n \in \mathbb{N}$  e considere o espaço  $X = \mathbb{R}^{n \times n}$  das matrizes reais  $n \times n$ . Considere os subespaços

$$S_1 = \{A \in X; A^{\text{tr}} = A\} \text{ (simétricas),}$$

$$S_2 = \{A \in X; A = (a_{ij})_{i,j=1}^n, a_{ij} = 0, \text{ para } |i - j| > 1\} \text{ (tridiagonais)}$$

$$S_3 = \{A \in X; A = (a_{ij})_{i,j=1}^n, a_{ij} = a_{ji}, i, j = 1, \dots, n, \\ a_{ij} = 0, \text{ para } |i - j| > 1\} \text{ (simétricas tridiagonais)}$$

Ache as dimensões de cada um desses subespaços e encontre uma base para  $S_3$ .

**5.41.** Prove o resultado descrito na [Observação 4.1](#) de existência e unicidade de uma função multilinear alternante  $d_{\mathcal{b}} : X^n \rightarrow \mathbb{K}$ , em um espaço vetorial  $X$  de dimensão finita  $n \in \mathbb{N}$  sobre um corpo  $\mathbb{K}$ , normalizado para valer  $d_{\mathcal{b}}(\mathbf{w}_1, \dots, \mathbf{w}_n) = 1$ , em uma determinada base  $\mathcal{b} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ .

**5.42.** Em relação ao funcional multilinear discutido no [Exercício 5.41](#), mostre que se  $\mathcal{a} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  e  $\mathcal{b} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  são duas bases de  $X$ , então

$$d_{\mathcal{a}}(\mathbf{u}_1, \dots, \mathbf{u}_n) = \frac{d_{\mathcal{a}}(\mathbf{v}_1, \dots, \mathbf{v}_n)}{d_{\mathcal{b}}(\mathbf{w}_1, \dots, \mathbf{w}_n)} d_{\mathcal{b}}(\mathbf{u}_1, \dots, \mathbf{u}_n),$$

para vetores  $\mathbf{u}_1, \dots, \mathbf{u}_n \in X$  quaisquer.



## CAPÍTULO 2

### Funcionais Lineares e Dualidade

#### 1. Funcionais lineares e o dual algébrico

**1.1. Definição.** Um **funcional** em um espaço vetorial  $X$  sobre um corpo  $\mathbb{K}$  é qualquer função  $f : X \rightarrow \mathbb{K}$  definida em  $X$  e com valores escalares em  $\mathbb{K}$ . Por exemplo, o comprimento de arco

$$\ell(f) = \int_I \sqrt{1 + |f'(x)|^2} \, dx$$

do gráfico  $y = f(x)$ ,  $x \in I$ , de uma curva suave  $f : I \rightarrow \mathbb{R}$  sobre um intervalo limitado  $I \subset \mathbb{R}$  pode ser visto como um funcional (não linear) no espaço vetorial real  $X = \mathcal{C}^1(I)$ , sobre  $\mathbb{K} = \mathbb{R}$ .

De particular interesse em Álgebra Linear, no entanto, são os **funcionais lineares**, também chamadas de **formas lineares**, i.e. que satisfazem

$$f(\lambda \mathbf{u} + \mathbf{v}) = \lambda f(\mathbf{u}) + f(\mathbf{v}),$$

para todo  $\mathbf{u}, \mathbf{v} \in X$  e  $\lambda \in \mathbb{K}$ .

O conjunto dos funcionais lineares forma um outro espaço vetorial chamado de *dual*, ou *dual algébrico*. Nesse sentido, temos a seguinte definição.

**DEFINIÇÃO 1.1 (Dual algébrico).** *Dado um espaço vetorial  $X$  sobre um corpo  $\mathbb{K}$ , o **dual algébrico**  $X^*$ , ou simplesmente **dual** no contexto de dimensão finita, é o conjunto dos **funcionais lineares** em  $X$ ,*

$$X^* = \{f : X \rightarrow \mathbb{K}; f(\lambda \mathbf{u} + \mathbf{v}) = \lambda f(\mathbf{u}) + f(\mathbf{v}), \forall \mathbf{u}, \mathbf{v} \in X, \lambda \in \mathbb{K}\}$$

*O dual algébrico  $X^*$  é um espaço vetorial quando munido das operações de adição e multiplicação por escalar definidas pontualmente, i.e. para todo  $f, g \in X^*$ , a adição  $f + g \in X^*$  é dada por*

$$(f + g)(\mathbf{u}) = f(\mathbf{u}) + g(\mathbf{u}), \quad \forall \mathbf{u} \in X,$$

*e, para todo  $\lambda \in \mathbb{K}$  e todo  $f \in X^*$ , a multiplicação por escalar  $\lambda f \in X^*$  é dada por*

$$(\lambda f)(\mathbf{u}) = \lambda f(\mathbf{u}), \quad \forall \mathbf{u} \in X.$$

OBSERVAÇÃO 1.1. Em dimensão infinita, consideramos, em muitos casos, espaços vetoriais munidos de alguma topologia e tais que as operações de adição e multiplicação por escalar sejam contínuas nessa topologia. A partir disso, podemos considerar funcionais  $f : X \rightarrow \mathbb{K}$  que sejam contínuos, usando a topologia de  $X$ . Em particular, podemos restringir os funcionais lineares aos que sejam contínuos, formando o que chamamos de **dual topológico**  $X' = \{f : X \rightarrow \mathbb{K}; f \text{ linear e contínuo}\}$ . Nesse contexto, o dual topológico é chamado simplesmente de dual. No caso de  $X$  ser de dimensão finita e munido de uma topologia compatível com as operações vetoriais, todo funcional linear é contínuo, de modo que as duas definições de espaço dual coincidem. Mas em dimensão infinita, temos, em geral, apenas  $X' \subset X^*$ .

### 1.2. Exemplos de funcionais lineares.

EXEMPLO 1.1 (Produto matriz-vetor com uma matriz de uma linha). Em  $X = \mathbb{R}^n$ ,  $n \in \mathbb{N}$ , dado  $\mathbf{a} = [a_1 \ \cdots \ a_n] \in \mathbb{R}^{1 \times n}$ , a função  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  dada por

$$f(\mathbf{u}) = \mathbf{a}\mathbf{u} = [a_1 \ \cdots \ a_n] \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = a_1x_1 + \cdots + a_nx_n$$

para todo  $\mathbf{u} = (x_1, \dots, x_n) \in \mathbb{R}^n$ , é um funcional linear em  $\mathbb{R}^n$ .

EXEMPLO 1.2 (Produto escalar em  $\mathbb{R}^n$  com um vetor dado). Em  $X = \mathbb{R}^n$ ,  $n \in \mathbb{N}$ , dado  $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{R}^n$ , a função  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  definida por

$$f(\mathbf{u}) = \mathbf{a} \cdot \mathbf{u} = a_1x_1 + \cdots + a_nx_n,$$

para todo  $\mathbf{u} = (x_1, \dots, x_n) \in \mathbb{R}^n$ , é um funcional linear em  $\mathbb{R}^n$ .

EXEMPLO 1.3 (Partes real e imaginária de um complexo). Olhando  $X = \mathbb{C}$  como espaço vetorial sobre  $\mathbb{K} = \mathbb{R}$ , as funções  $f : \mathbb{C} \rightarrow \mathbb{R}$  dadas por  $f(z) = f(x + iy) = x$  e  $f(z) = f(x + iy) = y$  são funcionais lineares em  $X$ . Escrevendo  $f(z) = f(x + iy) = x + i0$  e  $f(z) = f(x + iy) = y + i0$ , podemos interpretar essas funções, também, como funcionais lineares em  $\mathbb{C}$  como espaço vetorial sobre si mesmo,  $\mathbb{K} = \mathbb{C}$ .

EXEMPLO 1.4 (Valor de uma função em um ponto). Considerando o espaço  $X = \mathbb{R}^I$  de funções de um intervalo  $I \subset \mathbb{R}$  em  $\mathbb{R}$ , dado  $x_0 \in I$ , a função  $f : \mathbb{R}^I \rightarrow \mathbb{R}$  definida por

$$f(u) = u(x_0)$$

é um funcional linear em  $X$ .

EXEMPLO 1.5 (Valor médio). Considerando o espaço  $X = \mathcal{C}(I, \mathbb{R})$  em um intervalo  $I \subset \mathbb{R}$ , a função

$$f(u) = \frac{1}{|I|} \int_I u(x) \, dx,$$

onde  $|I|$  é o comprimento do intervalo, define um funcional linear em  $X$ .

EXEMPLO 1.6 (Derivada em um ponto). Considerando o espaço  $X = \mathcal{C}^1(I, \mathbb{R})$  em um intervalo  $I \subset \mathbb{R}$  e fixando  $x_0 \in I$ , a função

$$f(u) = u'(x_0)$$

define um funcional linear em  $X$ .

EXEMPLO 1.7 (Traço de uma matriz). Considerando o espaço  $X = \mathbb{C}^{n \times n}$  das matrizes complexas quadradas  $A = (a_{ij})_{i,j=1}^n$ , o traço

$$\text{tr}(A) = a_{11} + \cdots + a_{nn}$$

define um funcional linear em  $X$ .

EXEMPLO 1.8 (Diferencial). Um exemplo fundamental é o de diferencial de um funcional (usualmente não linear), quando o espaço vetorial está munido de uma norma. Mais precisamente, se  $\mathbf{f} : X \rightarrow \mathbb{K}$  é um funcional em um espaço vetorial normado sobre  $\mathbb{K} = \mathbb{R}$  ou  $\mathbb{C}$ , com a norma denotada por  $\|\cdot\|$ , dizemos que  $\mathbf{f}$  é **diferenciável** em um ponto  $\mathbf{u} \in X$ , quando existe um funcional linear  $\mathbf{g} : X \rightarrow Y$ , que é a **diferencial** de  $\mathbf{f}$  nesse ponto e é denotada por  $d\mathbf{f}(\mathbf{u}) = \mathbf{g}$ , que satisfaz

$$\frac{|\mathbf{f}(\mathbf{u} + \mathbf{h}) - \mathbf{f}(\mathbf{u}) - (d\mathbf{f}(\mathbf{u}))(\mathbf{h})|}{\|\mathbf{h}\|} \rightarrow 0, \quad \text{quando } \|\mathbf{h}\| \rightarrow \mathbf{0}, \mathbf{h} \neq \mathbf{0}.$$

Quando  $\mathbf{u} = (x_1, \dots, x_n) \in X = \mathbb{R}^n$  e  $\mathbf{f}(\mathbf{u}) = f(x_1, \dots, x_n)$ ,  $n \in \mathbb{N}$ , então a diferencial calculada em um vetor  $\mathbf{h}$  pode ser *representada* por uma multiplicação por escalar  $d\mathbf{f}(\mathbf{u}) \cdot \mathbf{h}$ , onde  $d\mathbf{f}(\mathbf{u}) = (\partial_{x_1} f(x_1, \dots, x_n), \partial_{x_n} f(x_1, \dots, x_n)) \in \mathbb{R}^n$  é o chamado **vetor gradiente** de  $\mathbf{f}$ .

OBSERVAÇÃO 1.2 (Aplicações). Como veremos a seguir, funcionais lineares podem ser usados para representar hiperplanos. Em particular, estão associados a hiperplanos tangentes ao gráfico de funcionais não-lineares. Também são usados para definir sub-diferenciais de funções convexas. Podem, ainda, ser usados para definir polítopos e poliedros; separar conjuntos convexos; classificar dados via máquinas de vetores de suporte (em inglês *support vector machine* - SVM); etc.

Classificação de dados via hiperplano afim

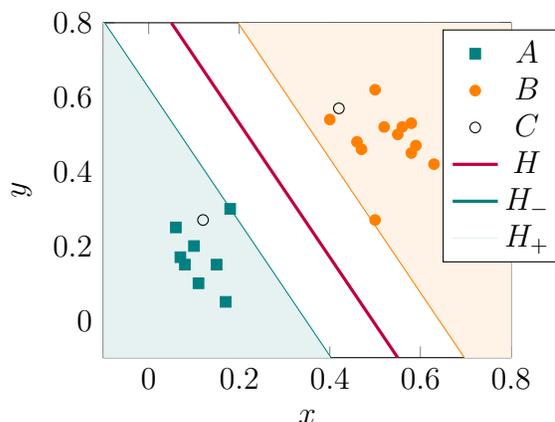


FIGURA 1.1. Classificação de dados via hiperplano. Dois grupos de dados já classificados, divididos em grupos  $A$  e  $B$ , são usados para treinar o método, gerando um hiperplano  $H = \{f = \eta\}$  que separa os grupos em semi-hiperplanos, e.g.  $A$  em  $H_- = \{f \leq \eta - \delta\}$  e  $B$  em  $H_+ = \{f \geq \eta + \delta\}$ . Uma vez treinado, o hiperplano pode ser usado para classificar novos dados ( $C$ ). Os vetores nos bordos  $\{f = \eta - \delta\}$  e  $\{f = \eta + \delta\}$  são os *vetores de suporte*, no método de classificação conhecido como *máquina de vetores de suporte*, em inglês *support vector machine (SVM)*.

**1.3. Produto de dualidade.** Consideramos, inicialmente, a ação  $\mathbf{u} \mapsto \mathbf{f}(\mathbf{u})$ , em  $X$ , de um dado funcional linear  $\mathbf{f} \in X^*$ . Em seguida, consideramos o espaço de todos os funcionais lineares e definimos adição e multiplicação por escalar entre eles, o que significa considerar a ação  $\mathbf{f} \mapsto \mathbf{f}(\mathbf{u})$ , para funcionais lineares em  $X^*$ , para um dado  $\mathbf{u} \in X$ . Podemos combinar as duas ações em um mapeamento  $(\mathbf{f}, \mathbf{u}) \mapsto \mathbf{f}(\mathbf{u})$  e definir um funcional no produto cartesiano  $X^* \times X$ . Como esse funcional é linear tanto em  $\mathbf{f}$  como em  $\mathbf{u}$ , dizemos que ele é um **funcional bilinear**, ou uma **forma bilinear**. E por ser bilinear, é natural usamos uma notação que torne isso mais natural, e.g. de uma forma bilinear  $\langle \cdot, \cdot \rangle$  definida no produto cartesiano  $X^* \times X$ . Vamos formalizar isso e ver algumas de suas propriedades.

DEFINIÇÃO 1.2 (Produto de dualidade). *Dado um espaço vetorial  $X$  sobre um corpo  $\mathbb{K}$ , definimos o **produto de dualidade** entre o seu dual e o espaço como sendo*

o funcional bilinear  $\langle \cdot, \cdot \rangle : X^* \times X \rightarrow \mathbb{K}$  dado por

$$\langle \mathbf{f}, \mathbf{u} \rangle = \mathbf{f}(\mathbf{u}),$$

para  $\mathbf{f} \in X^*$  e  $\mathbf{u} \in X$ .

A bilinearidade se escreve

$$\langle \mathbf{f}, \mathbf{u} + \lambda \mathbf{v} \rangle = \mathbf{f}(\mathbf{u} + \lambda \mathbf{v}) = \mathbf{f}(\mathbf{u}) + \lambda \mathbf{f}(\mathbf{v}) = \langle \mathbf{f}, \mathbf{u} \rangle + \lambda \langle \mathbf{f}, \mathbf{v} \rangle$$

e

$$\langle \mathbf{f} + \alpha \mathbf{g}, \mathbf{u} \rangle = (\mathbf{f} + \alpha \mathbf{g})(\mathbf{u}) = \mathbf{f}(\mathbf{u}) + \alpha \mathbf{g}(\mathbf{u}) = \langle \mathbf{f}, \mathbf{u} \rangle + \alpha \langle \mathbf{g}, \mathbf{u} \rangle.$$

OBSERVAÇÃO 1.3. Em certos casos, quando trabalhando com mais de um espaço vetorial, é útil qualificar os espaços em cada produto de dualidade, escrevendo, por exemplo,

$$\langle \mathbf{f}, \mathbf{u} \rangle_{X^*, X}.$$

Observe que a ordem dos espaços acompanha a ordem dos argumentos,  $\mathbf{f} \in X^*$  primeiro e  $\mathbf{u} \in X$  em seguida.

## 2. Núcleo de um funcional linear e hiperplanos

### 2.1. O núcleo como um subespaço.

PROPOSIÇÃO 2.1. *Dado um funcional linear  $\mathbf{f} \in X^*$  em um espaço vetorial  $X$  sobre um corpo  $\mathbb{K}$ , o **núcleo** de  $\mathbf{f}$ , definido por*

$$\ker(\mathbf{f}) = \{\mathbf{u} \in X; \langle \mathbf{f}, \mathbf{u} \rangle = 0\},$$

é um subespaço de  $X$ .

DEMONSTRAÇÃO. Primeiramente, vamos verificar que  $\mathbf{0} \in \ker(\mathbf{f})$ , que é condição necessária para ser subespaço. Como  $\mathbf{0}$  é o elemento neutro aditivo e  $\mathbf{f}$  é linear, temos

$$\langle \mathbf{f}, \mathbf{0} \rangle = \langle \mathbf{f}, \mathbf{0} + \mathbf{0} \rangle = \langle \mathbf{f}, \mathbf{0} \rangle + \langle \mathbf{f}, \mathbf{0} \rangle = 2\langle \mathbf{f}, \mathbf{0} \rangle.$$

A única possibilidade disso ser verdade em  $\mathbb{K}$  é com  $\langle \mathbf{f}, \mathbf{0} \rangle = 0$ , de modo que  $\mathbf{0} \in \ker(\mathbf{f})$ .

Agora, basta observar que, pela linearidade de  $\mathbf{f}$ , se  $\mathbf{u}, \mathbf{v} \in \ker(\mathbf{f})$  e  $\lambda \in \mathbb{K}$ , então

$$\langle \mathbf{f}, \mathbf{u} + \mathbf{v} \rangle = \langle \mathbf{f}, \mathbf{u} \rangle + \langle \mathbf{f}, \mathbf{v} \rangle = 0 + 0 = 0,$$

e

$$\langle \mathbf{f}, \lambda \mathbf{u} \rangle = \lambda \langle \mathbf{f}, \mathbf{u} \rangle = \lambda 0 = 0,$$

de modo que  $\mathbf{u} + \mathbf{v} \in \ker(\mathbf{f})$  e  $\lambda \mathbf{u} \in \ker(\mathbf{f})$ , mostrando que  $\ker(\mathbf{f})$  é fechado para as operações de adição e multiplicação por escalar.  $\square$

## 2.2. Exemplos de núcleos.

EXEMPLO 2.1. Em  $X = \mathbb{R}^2$ , dados  $\mathbf{a} = (a, b) \in \mathbb{R}^2$  e o funcional linear  $\langle \mathbf{f}, \mathbf{u} \rangle = \mathbf{a} \cdot \mathbf{u} = ax + by$ , para  $\mathbf{u} = (x, y)$ , então o núcleo de  $\mathbf{f}$  é uma reta passando pela origem:

$$\ker(\mathbf{f}) = \{(x, y) \in \mathbb{R}^2; ax + by = 0\}$$

EXEMPLO 2.2 (Partes real e imaginária de um complexo). Olhando  $X = \mathbb{C}$  como espaço vetorial sobre  $\mathbb{K} = \mathbb{R}$  e considerando  $\mathbf{f}(z) = x$ , para  $z = x + iy \in \mathbb{C}$ , o núcleo de  $\mathbf{f}$  é o eixo imaginário.

EXEMPLO 2.3. Seja  $X = \mathcal{P}(\mathbb{R})$  o espaço dos polinômios na reta real. Considere o funcional linear

$$\mathbf{f}(p) = p(0).$$

O núcleo de  $\mathbf{f}$  é composto por todos os polinômios que se anulam na origem, ou seja

$$\ker(\mathbf{f}) = \{p \in \mathcal{P}(\mathbb{R}); p(0) = 0\} = \{p; p(x) = xq(x), q \in \mathcal{P}(\mathbb{R})\} = x\mathcal{P}(\mathbb{R}).$$

EXEMPLO 2.4 (Valor médio). Considerando o espaço  $X = \mathcal{C}(I, \mathbb{R})$  em um intervalo limitado  $I \subset \mathbb{R}$  e o funcional linear

$$\mathbf{f}(u) = \frac{1}{|I|} \int_I u(x) dx,$$

onde  $|I|$  é o comprimento do intervalo, o núcleo de  $\mathbf{f}$  é o subespaço das funções de média zero:

$$\ker(\mathbf{f}) = \left\{ u \in \mathcal{C}(I, \mathbb{R}); \int_I u(x) dx = 0 \right\}.$$

EXEMPLO 2.5 (Derivada em um ponto). Considerando o espaço  $X = \mathcal{C}^1(I, \mathbb{R})$  em um intervalo  $I \subset \mathbb{R}$  e fixando  $x_0 \in I$ , o núcleo do funcional

$$\mathbf{f}(u) = u'(x_0)$$

é o conjunto das funções continuamente diferenciáveis com um ponto crítico em  $x = x_0$ .

EXEMPLO 2.6 (Traço de uma matriz). Considerando o espaço  $X = \mathbb{C}^{n \times n}$  das matrizes complexas quadradas  $A = (a_{ij})_{i,j=1}^n$ , o núcleo do traço

$$\text{tr}(A) = a_{11} + \cdots + a_{nn}$$

é o subespaço das matrizes com traço zero, que caracteriza as matrizes  $A$  associadas aos sistemas  $\dot{\mathbf{u}} = A\mathbf{u}$  que preservam volume no espaço de fase.

### 2.3. Co-dimensão do núcleo de um funcional não nulo.

PROPOSIÇÃO 2.2. *Seja  $X$  um espaço vetorial e seja  $f \in X^*$  um funcional linear não nulo. Então  $\ker(f)$  tem co-dimensão um, ou seja, é um hiperplano.*

DEMONSTRAÇÃO. Basta mostrar que  $\ker(f)$  tem um complementar de dimensão um, ou seja, que tem um complementar gerado por um único vetor não nulo. Para tanto, deve ser um vetor fora do núcleo de  $f$ . Seja, então,  $\mathbf{w} \in X$  tal que  $\langle f, \mathbf{w} \rangle \neq 0$ , que existe já que  $f$  é não nulo. Basta, então, mostrar que  $X = \ker(f) \oplus \text{span}\{\mathbf{w}\}$ . Para isso, podemos mostrar que, para todo  $\mathbf{u} \in X$ , existe um único  $\mathbf{v} \in \ker(f)$  e um único  $\lambda \in \mathbb{K}$  tais que

$$\mathbf{u} = \mathbf{v} + \lambda \mathbf{w}.$$

Seja, então,  $\mathbf{u} \in X$  qualquer. Tomemos

$$\lambda = \frac{\langle f, \mathbf{u} \rangle}{\langle f, \mathbf{w} \rangle}$$

e

$$\mathbf{v} = \mathbf{u} - \lambda \mathbf{w}.$$

Assim,

$$\mathbf{u} = \lambda \mathbf{w} + \mathbf{v}$$

com

$$\langle f, \mathbf{v} \rangle = \langle f, \mathbf{u} - \lambda \mathbf{w} \rangle = \langle f, \mathbf{u} \rangle - \lambda \langle f, \mathbf{w} \rangle = 0,$$

de modo que  $\mathbf{v} \in \ker(f)$ , nos dando a representação desejada. Para verificar a unicidade dessa representação, sejam  $\lambda_1, \lambda_2 \in \mathbb{K}$  e  $\mathbf{v}_1, \mathbf{v}_2 \in \ker(f)$  tais que

$$\lambda_1 \mathbf{w} + \mathbf{v}_1 = \mathbf{u} = \lambda_2 \mathbf{w} + \mathbf{v}_2.$$

Então, como  $\langle f, \mathbf{v}_1 \rangle = \langle f, \mathbf{v}_2 \rangle = 0$  e  $f$  é linear, temos que

$$\begin{aligned} \lambda_1 \langle f, \mathbf{w} \rangle &= \lambda_1 \langle f, \mathbf{w} \rangle + \langle f, \mathbf{v}_1 \rangle = \langle f, \lambda_1 \mathbf{w} + \mathbf{v}_1 \rangle \\ &= \langle f, \lambda_2 \mathbf{w} + \mathbf{v}_2 \rangle = \lambda_2 \langle f, \mathbf{w} \rangle + \langle f, \mathbf{v}_2 \rangle = \lambda_2 \langle f, \mathbf{w} \rangle. \end{aligned}$$

Como  $\langle f, \mathbf{w} \rangle \neq 0$ , segue que

$$\lambda_1 = \lambda_2.$$

Assim,

$$\mathbf{v}_2 - \mathbf{v}_1 = \lambda_1 \mathbf{w} - \lambda_2 \mathbf{w} = 0,$$

de maneira que

$$\mathbf{v}_1 = \mathbf{v}_2,$$

completando a demonstração da unicidade. Com essa representação única, deduzimos que  $X = \ker(f) \oplus \text{span}\{\mathbf{w}\}$ . Logo,  $\ker(f)$  tem co-dimensão um.  $\square$

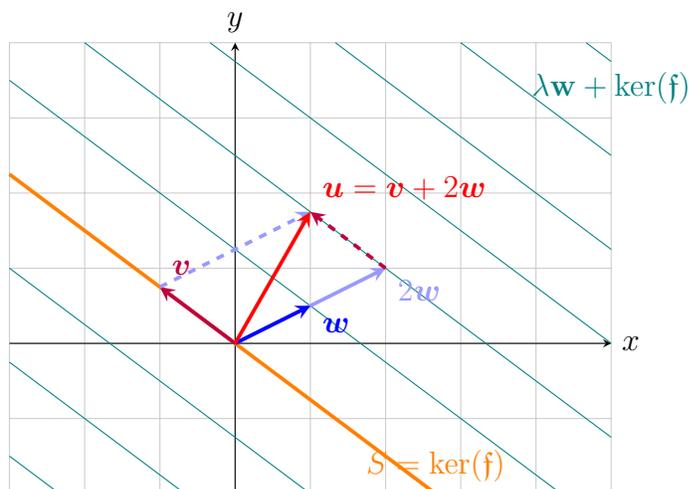


FIGURA 2.1. Ilustração do núcleo  $S = \ker(f)$  de um funcional  $f \in X^*$  não nulo; do conjunto de hiperplanos afim  $\lambda \mathbf{w} + \ker(f)$ , que forma o espaço quociente  $X/\ker(f)$ ; e da representação  $\mathbf{u} = \mathbf{v} + 2\mathbf{w}$  de um vetor  $\mathbf{u}$  em  $X = \ker(f) \oplus \text{span } \mathbf{w}$ . A escolha de  $\mathbf{w} \in X \setminus \ker(f)$  é como uma escolha de *escala*, ou *unidade de medida*, para o espaço complementar ao núcleo.

O seguinte resultado segue da proposição acima.

**COROLÁRIO 2.1.** *Dado um funcional linear  $f \in X^*$  não-nulo em um espaço vetorial  $X$  sobre um corpo  $\mathbb{K}$  e dado  $\mathbf{w} \in X$  tal que  $\langle f, \mathbf{w} \rangle \neq 0$ , temos que, para qualquer  $\mathbf{u} \in X$ , existem únicos  $\lambda \in \mathbb{K}$  e  $\mathbf{v} \in \ker(f)$  tais que*

$$\mathbf{u} = \lambda \mathbf{w} + \mathbf{v}.$$

**2.4. Espaço quociente ao núcleo.** Como  $X = \ker(f) \oplus \text{span}\{\mathbf{w}\}$ , no caso de um  $f \in X^*$  não nulo, segue, da **Proposição 2.7**, que  $X/\ker(f)$  é isomorfo a  $\text{span}\{\mathbf{w}\}$  e, assim, tem dimensão um. Isso nos dá o seguinte corolário.

**COROLÁRIO 2.2.** *Seja  $X$  um espaço vetorial de dimensão finita e seja  $f \in X^*$  um funcional linear não nulo. Então  $X/\ker(f)$  tem dimensão um, sendo isomorfo a  $\text{span}\{\mathbf{w}\}$ , para qualquer  $\mathbf{w} \in X \setminus \ker(f)$ .*

**OBSERVAÇÃO 2.1.** Vale lembrar, da **Definição 2.20**, que

$$X/\ker(f) = \{\mathbf{w} + \ker(f); \mathbf{w} \in X\}.$$

Como  $\ker(\mathbf{f})$  tem co-dimensão um, os conjuntos  $\mathbf{w} + \ker(\mathbf{f})$  são hiperplanos afim, de modo que o espaço quociente é a união de todos os hiperplanos afim paralelos a  $\ker(\mathbf{f})$ .

**OBSERVAÇÃO 2.2.** Uma equação linear  $u' = Au$ , onde  $u = u(t) \in \mathbb{R}^n$ ,  $A \in \mathbb{R}^{n \times n}$ ,  $n \in \mathbb{N}$ , e  $u' = u'(t)$  é a derivada (temporal) em  $t$ , tem solução  $u(t) = e^{tA}u_0$ ,  $t \in \mathbb{R}$ . A família de operadores-solução  $e^{tA} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  age no espaço de fases  $\mathbb{R}^n$ . Essa família pode aumentar, diminuir ou preservar volumes em  $\mathbb{R}^n$  de acordo com o determinante de  $e^{tA}$ . Esse determinante está associado ao traço de  $A$  por  $\det(e^{tA}) = e^{t\text{tr}(A)}$ . Em particular, matrizes de traço nulo geram operadores-solução que preservam volume. O subespaço  $S$  de matrizes de traço nulo separa o espaço das matrizes em dois conjuntos conexos, um formado pelas matrizes que geram operadores-solução *expansores* e outro formado pelas matrizes que geram operadores-solução *contratores*.

**2.5. Caracterização de hiperplanos.** Vimos, então, que o núcleo de um funcional linear não nulo em um espaço de dimensão finita é um hiperplano, ou seja, tem co-dimensão um. Em outras palavras, partimos de um funcional não nulo e chegamos em um núcleo que é um hiperplano.

O resultado a seguir complementa o resultado acima, partindo de um hiperplano, i.e. de um subespaço linear de co-dimensão um, e chegando em um funcional linear com esse subespaço como núcleo. Isso nos dá, de fato, uma caracterização dos hiperplanos como núcleos de funcionais lineares.

**PROPOSIÇÃO 2.3.** *Seja  $X$  um espaço vetorial de dimensão finita. Dados um subespaço vetorial  $S \subset X$  de co-dimensão um e um vetor  $\mathbf{w} \in X \setminus S$ , existe um único funcional linear  $\mathbf{f} \in X^*$  tal que  $S = \ker(\mathbf{f})$  e  $\mathbf{f}(\mathbf{w}) = 1$ .*

**DEMONSTRAÇÃO.** Como  $S$  é de co-dimensão um e  $\mathbf{w} \notin S$ , então  $X = S \oplus \text{span}\{\mathbf{w}\}$ . Assim, todo  $\mathbf{u} \in X$  pode ser escrito de maneira única como soma de um elemento de  $S$  e de um múltiplo de  $\mathbf{w}$ , i.e. existem  $\mathbf{v} \in S$  e  $\lambda \in \mathbb{K}$  únicos tais que

$$\mathbf{u} = \mathbf{v} + \lambda\mathbf{w}.$$

Como essa decomposição é única, podemos definir, sem ambiguidade, um funcional  $\mathbf{f} : X \rightarrow \mathbb{K}$  dado por

$$\langle \mathbf{f}, \mathbf{u} \rangle = \lambda.$$

Sendo única, podemos ver que essa decomposição é uma operação linear, i.e. se  $\mathbf{u}_1 = \mathbf{v}_1 + \lambda_1\mathbf{w}$  e  $\mathbf{u}_2 = \mathbf{v}_2 + \lambda_2\mathbf{w}$ , então

$$\mathbf{u}_1 + \alpha\mathbf{u}_2 = \mathbf{v}_1 + \alpha\mathbf{v}_2 + (\lambda_1 + \alpha\lambda_2)\mathbf{w},$$

com  $\mathbf{v}_1 + \alpha\mathbf{v}_2 \in S$  e  $\lambda_1 + \alpha\lambda_2 \in \mathbb{K}$ . Dessa forma,

$$\langle \mathbf{f}, \mathbf{u}_1 + \alpha\mathbf{u}_2 \rangle = \lambda_1 + \alpha\lambda_2 = \langle \mathbf{f}, \mathbf{u}_1 \rangle + \alpha\langle \mathbf{f}, \mathbf{u}_2 \rangle,$$

mostrando que  $f$  é linear em  $X$ , i.e.  $f \in X^*$ .

Como  $\mathbf{w} = \mathbf{0} + 1\mathbf{w}$ , com  $\mathbf{0} \in S$ , temos, por construção, que

$$\langle f, \mathbf{w} \rangle = 1.$$

Assim, construímos  $f \in X^*$  com  $\langle f, \mathbf{w} \rangle = 1$  e  $\ker(f) = S$ . Essa  $f$  é única. De fato, se  $\tilde{f} \in X^*$  se anula em  $S$  e vale 1 em  $\mathbf{w}$ , então, usando a decomposição acima e a linearidade de  $\tilde{f}$ ,

$$\langle \tilde{f}, \mathbf{u} \rangle = \langle \tilde{f}, \mathbf{v} + \lambda\mathbf{w} \rangle = \langle \tilde{f}, \mathbf{v} \rangle + \lambda\langle \tilde{f}, \mathbf{w} \rangle = 0 + \lambda = \lambda = \langle f, \mathbf{u} \rangle,$$

para todo  $\mathbf{u} \in X$ . Ou seja,  $\tilde{f} = f$ , nos dando a unicidade.  $\square$

**OBSERVAÇÃO 2.3.** A unicidade na **Proposição 2.3** é dada uma vez que fixamos  $\mathbf{w} \in X \setminus \ker(f)$  e forçamos que  $\langle f, \mathbf{w} \rangle = 1$ . Essa escolha é como uma escolha de *escala*, ou *unidade de medida*, para o espaço complementar ao núcleo. Podemos mudar  $\mathbf{w}$  ou mudar o valor de  $\langle f, \mathbf{w} \rangle$  e termos diversas representações diferentes, em diferentes escalas. Mas uma vez fixada a escala, só há uma escolha possível. Dois  $\mathbf{w}_1, \mathbf{w}_2$  que estejam no mesmo hiperplano afim paralelo a  $S$  determinam o mesmo funcional linear e dão uma mesma escala para  $X/S$ . Veja **Seção 2.4** e **Seção 5.1**.

**2.6. Propriedades úteis.** Os seguinte corolários são, por vezes, úteis (veja, e.g. **Teorema 4.1**).

**COROLÁRIO 2.3.** *Seja  $X$  um espaço vetorial de dimensão finita. Dado  $\mathbf{w} \in X$  não nulo, existe  $f \in X^*$  tal que  $\langle f, \mathbf{w} \rangle = 1$ .*

**DEMONSTRAÇÃO.** Basta considerar um espaço complementar  $S$  qualquer do subespaço  $\text{span}\{\mathbf{w}\}$  e aplicar o **Proposição 2.3**.  $\square$

**OBSERVAÇÃO 2.4.** Observe que  $f$  no **Corolário 2.3** pode não ser único, já que a escolha do complementar  $S$  é arbitrária. Só é única no caso em que  $X$  é unidimensional.

**OBSERVAÇÃO 2.5.** O resultado do **Corolário 2.3** continua válido no caso de dimensão infinita, mas a demonstração requer o uso de base de Hamel. Mais importante do que isso é que, em alguns casos, o tal funcional  $f \in X^*$  pode não ser contínuo, o que é relevante no caso de dimensão infinita (em dimensão finita, como falamos na **Observação 1.1**, todos os funcionais lineares são contínuos). O lado positivo disso é que se todos os funcionais lineares contínuos se anularem em um determinado subespaço, então esse subespaço é *denso* no espaço todo. Em relação a isso, veja, por exemplo, **Exemplo 1.3**.

**COROLÁRIO 2.4.** *Seja  $X$  um espaço vetorial de dimensão finita e sejam  $\mathbf{u}, \mathbf{v} \in X$ . Se  $\langle f, \mathbf{u} \rangle = \langle f, \mathbf{v} \rangle$  para todo  $f \in X^*$  então  $\mathbf{u} = \mathbf{v}$ .*

DEMONSTRAÇÃO. Por linearidade, temos  $\langle \mathbf{f}, \mathbf{u} - \mathbf{v} \rangle = 0$ , para todo  $\mathbf{f} \in X^*$ . Se  $\mathbf{u} - \mathbf{v} \neq \mathbf{0}$ , então segue da [Corolário 2.3](#) que existe  $\mathbf{f} \in X^*$  tal que  $\langle \mathbf{f}, \mathbf{u} - \mathbf{v} \rangle \neq 0$ . Mas isso contradiz a hipótese. Logo,  $\mathbf{u} - \mathbf{v} = \mathbf{0}$ , ou seja,  $\mathbf{u} = \mathbf{v}$ .  $\square$

**2.7. Caracterização de hiperplanos afim.** A caracterização anterior de um subespaço de co-dimensão um como o núcleo de um funcional linear não nulo se estende facilmente para hiperplanos afim. Vimos que um hiperplano afim é um conjunto  $\mathbf{w} + S$  onde  $S$  é um subespaço de co-dimensão um. Vamos ver, abaixo, uma caracterização de um hiperplano afim como conjunto de nível de um funcional linear.

PROPOSIÇÃO 2.4. *Um hiperplano afim pode ser caracterizado como sendo um conjunto de nível de um funcional linear, i.e. um conjunto  $H \subset X$  é um hiperplano afim em um espaço vetorial  $X$  sobre um corpo  $\mathbb{K}$  se, e somente se, existe um funcional linear não-trivial  $\mathbf{f} \in X^*$  e um escalar  $\eta \in \mathbb{K}$  tal que*

$$H = \{\mathbf{f} = \eta\} = \{\mathbf{u} \in X; \langle \mathbf{f}, \mathbf{u} \rangle = \eta\}.$$

DEMONSTRAÇÃO. Seja  $H = \mathbf{w} + S$ , onde  $\mathbf{w} \in X$  e  $S$  é um hiperplano. Como  $S$  é um hiperplano,  $S$  tem co-dimensão um e existe um complementar  $S'$  de  $S$  que é de dimensão um. Nesse caso,  $S'$  tem uma base que é formada por um único elemento não nulo, digamos  $S' = \text{span}\{\mathbf{w}_1\}$ , para algum  $\mathbf{w}_1 \in X$ . Todo vetor  $\mathbf{u}$  em  $X$  pode ser decomposto de forma única como

$$\mathbf{u} = \mathbf{v} + \lambda \mathbf{w}_1,$$

onde  $\mathbf{v} \in S$  e  $\lambda \in \mathbb{K}$ . Como essa decomposição é única, podemos definir, sem ambiguidade, o funcional linear  $\mathbf{f}: X \rightarrow \mathbb{K}$  por

$$\langle \mathbf{f}, \mathbf{u} \rangle = \langle \mathbf{f}, \mathbf{v} + \lambda \mathbf{w}_1 \rangle = \lambda.$$

Em particular,  $\langle \mathbf{f}, \mathbf{u} \rangle = 0$  se, e somente se,  $\mathbf{u} \in S$ . Além disso,  $\langle \mathbf{f}, \mathbf{w}_1 \rangle = 1$ .

O funcional  $\mathbf{f}$  definido acima é, de fato, linear, pois se  $\mathbf{u}' = \mathbf{v}' + \lambda' \mathbf{w}_1$  é outro vetor em  $X$ , então

$$\langle \mathbf{f}, \mathbf{u} + \mathbf{u}' \rangle = \langle \mathbf{f}, \mathbf{v} + \mathbf{v}' + (\lambda + \lambda') \mathbf{w}_1 \rangle = \lambda + \lambda' = \langle \mathbf{f}, \mathbf{u} \rangle + \langle \mathbf{f}, \mathbf{u}' \rangle,$$

visto que  $\mathbf{v} + \mathbf{v}' \in S$ . Da mesma forma, dado  $\mu \in \mathbb{K}$ , temos  $\mu \mathbf{v} \in S$  de modo que

$$\langle \mathbf{f}, \mu \mathbf{u} \rangle = \langle \mathbf{f}, \mu \mathbf{v} + \lambda \mu \mathbf{w}_1 \rangle = \lambda \mu = \mu \langle \mathbf{f}, \mathbf{u} \rangle.$$

Isso mostra que  $\mathbf{f} \in X^*$ . Seja, agora,

$$\mu = \langle \mathbf{f}, \mathbf{w} \rangle.$$

Observe que, para  $\mathbf{u} = \mathbf{v} + \lambda \mathbf{w}_1$ , temos  $\mathbf{u} \in H$  se, e somente se,  $\mathbf{u} = \mathbf{w} + \tilde{\mathbf{v}}$ , para algum  $\tilde{\mathbf{v}} \in S$ . Nesse caso,

$$\mathbf{v} + \lambda \mathbf{w}_1 = \mathbf{w} + \tilde{\mathbf{v}},$$

de modo que

$$\lambda \mathbf{w}_1 - \mathbf{w} = \tilde{\mathbf{v}} - \mathbf{v} \in S.$$

Como  $\mathbf{f}$  só se anula em  $S$ , vemos que  $\mathbf{u} = \mathbf{v} + \lambda \mathbf{w}_1$  pertence a  $H$  se, e somente se,

$$\langle \mathbf{f}, \lambda \mathbf{w}_1 - \mathbf{w} \rangle = 0,$$

o que nos leva a

$$0 = \langle \mathbf{f}, \lambda \mathbf{w}_1 - \mathbf{w} \rangle = \lambda \langle \mathbf{f}, \mathbf{w}_1 \rangle - \langle \mathbf{f}, \mathbf{w} \rangle = \lambda - \mu,$$

ou seja,  $\mathbf{u} \in H$  se, e somente se,  $\langle \mathbf{f}, \mathbf{u} \rangle = \mu$ . Isso nos dá a caracterização  $H = \{\mathbf{f} = \mu\}$ , para essa  $\mathbf{f}$ .  $\square$

Essa caracterização pode ser usada para verificar se dois hiperplanos são paralelos, como mostra o resultado a seguir.

**PROPOSIÇÃO 2.5.** *Dois hiperplanos  $H_1$  e  $H_2$  em um espaço vetorial  $X$  sobre um corpo  $\mathbb{K}$  são paralelos quando existe um funcional linear não-trivial  $\mathbf{f} \in X^*$  e um escalar  $\beta \in \mathbb{K}$  tais que*

$$\mathbf{f}(\mathbf{u}_1 - \mathbf{u}_2) = \beta, \quad \forall \mathbf{u}_1 \in H_1, \forall \mathbf{u}_2 \in H_2. \quad (2.1)$$

**DEMONSTRAÇÃO.** A ideia é escrever  $H_1 = \mathbf{w}_1 + S$  e  $H_2 = \mathbf{w}_2 + S$ , tomar um funcional  $\mathbf{f} \in X^*$  com  $\ker(\mathbf{f}) = S$  e definir  $\beta = \langle \mathbf{f}, \mathbf{w}_2 \rangle - \langle \mathbf{f}, \mathbf{w}_1 \rangle$ . Em seguida, é só verificar que (2.1) vale. Deixamos os detalhes para o leitor ([Exercício 7.1](#)).  $\square$

**OBSERVAÇÃO 2.6.** Unindo a representação do espaço quociente  $X/\ker(\mathbf{f})$  como união dos hiperplanos afim paralelos ao núcleo do funcional linear dada na [Observação 2.1](#), i.e.  $X/\ker(\mathbf{f}) = \{\mathbf{u} + \ker(\mathbf{f})\}$ , com a caracterização de hiperplanos afim dada na [Proposição 2.4](#), podemos escrever

$$X/\ker(\mathbf{f}) = \{H_\eta, \eta \in \mathbb{K}\},$$

onde

$$H_\eta = \{\mathbf{f} = \eta\} = \{\mathbf{u} \in X; \langle \mathbf{f}, \mathbf{u} \rangle = \eta\}.$$

Essa caracterização nos dá, de fato, um isomorfismo  $\eta \mapsto H_\eta = \{\mathbf{f} = \eta\}$ , entre  $\mathbb{K}$  e  $X/\ker(\mathbf{f})$ .

### 3. Dimensão do espaço dual

A [Proposição 2.3](#) é fundamental para construirmos uma base para o dual  $X^*$ , como veremos agora.

**3.1. Funcionais que identificam os vetores de uma base.** Primeiramente, construímos funcionais que “enxergam” somente um dos vetores de uma dada base.

LEMA 3.1. *Seja  $X$  um espaço vetorial de dimensão  $n \in \mathbb{N}$  e seja  $\mathfrak{b} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  uma base de  $X$ . Para cada  $i = 1, \dots, n$ , existe um único funcional linear  $\mathfrak{f}_i \in X^*$  tal que*

$$\langle \mathfrak{f}_i, \mathbf{w}_j \rangle = \delta_{ij}, \quad i, j = 1, \dots, n,$$

onde  $\delta_{ij}$  é o Delta de Kronecker.

DEMONSTRAÇÃO. Basta aplicar a **Proposição 2.3** para cada  $i = 1, \dots, n$ , tomando  $\mathbf{w} = \mathbf{w}_i$  e  $S = S_i = \text{span}\{\mathbf{w}_j, j \neq i\}$ .  $\square$

**3.2. Base dual e dimensão.** Agora, vamos mostrar que esses funcionais formam uma base para o dual.

TEOREMA 3.1. *Sejam  $X$  um espaço vetorial de dimensão  $n \in \mathbb{N}$ ,  $\mathfrak{b} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  uma base de  $V$  e  $\mathfrak{f}_1, \dots, \mathfrak{f}_n \in X^*$  funcionais lineares tais que*

$$\langle \mathfrak{f}_i, \mathbf{w}_j \rangle = \delta_{ij}, \quad i, j = 1, \dots, n.$$

Então  $\mathfrak{b}^* = \{\mathfrak{f}_1, \dots, \mathfrak{f}_n\}$  é uma base de  $X^*$ .

DEMONSTRAÇÃO. Dado  $\mathbf{u} \in X$ , temos uma representação única em termos da base,

$$\mathbf{u} = \lambda_1 \mathbf{w}_1 + \dots + \lambda_n \mathbf{w}_n. \quad (3.1)$$

Com isso, observe que

$$\langle \mathfrak{f}_i, \mathbf{u} \rangle = \langle \mathfrak{f}_i, \lambda_1 \mathbf{w}_1 + \dots + \lambda_n \mathbf{w}_n \rangle = \lambda_1 \langle \mathfrak{f}_i, \mathbf{w}_1 \rangle + \dots + \lambda_n \langle \mathfrak{f}_i, \mathbf{w}_n \rangle = \lambda_i. \quad (3.2)$$

Agora, seja  $\mathfrak{f} \in X^*$  arbitrário. Nesse caso, temos, para  $\mathbf{u} \in X$  qualquer,

$$\begin{aligned} \langle \mathfrak{f}, \mathbf{u} \rangle &= \langle \mathfrak{f}, \lambda_1 \mathbf{w}_1 + \dots + \lambda_n \mathbf{w}_n \rangle \\ &= \lambda_1 \langle \mathfrak{f}, \mathbf{w}_1 \rangle + \dots + \lambda_n \langle \mathfrak{f}, \mathbf{w}_n \rangle \\ &= \langle \mathfrak{f}_1, \mathbf{u} \rangle \langle \mathfrak{f}, \mathbf{w}_1 \rangle + \dots + \langle \mathfrak{f}_n, \mathbf{u} \rangle \langle \mathfrak{f}, \mathbf{w}_n \rangle \\ &= \langle \mathfrak{f}, \mathbf{w}_1 \rangle \langle \mathfrak{f}_1, \mathbf{u} \rangle + \dots + \langle \mathfrak{f}, \mathbf{w}_n \rangle \langle \mathfrak{f}_n, \mathbf{u} \rangle \\ &= \langle \mathfrak{f}(\mathbf{w}_1) \mathfrak{f}_1 + \dots + \mathfrak{f}(\mathbf{w}_n) \mathfrak{f}_n, \mathbf{u} \rangle, \end{aligned} \quad (3.3)$$

ou seja,

$$\mathfrak{f} = \alpha_1 \mathfrak{f}_1 + \dots + \alpha_n \mathfrak{f}_n \in \text{span}\{\mathfrak{f}_1, \dots, \mathfrak{f}_n\}, \quad (3.4)$$

com

$$\alpha_i = \langle \mathfrak{f}, \mathbf{w}_i \rangle, \quad i = 1, \dots, n.$$

Como isso vale para qualquer  $\mathfrak{f}$ , segue que  $\mathfrak{b}^*$  gera  $V^*$ .

Falta mostrar que  $\mathfrak{b}^*$  é linearmente independente. Suponha, então, que

$$\lambda_1 \mathfrak{f}_1 + \cdots + \lambda_n \mathfrak{f}_n = \mathfrak{o},$$

onde o elemento  $\mathfrak{o}$  do lado direito é o zero de  $X^*$ , ou seja, é o funcional nulo  $\mathfrak{o}(\mathbf{u}) = 0 \in \mathbb{K}$ , para todo  $\mathbf{u} \in X$ . Mais explicitamente, isso significa que

$$\lambda_1 \langle \mathfrak{f}_1, \mathbf{u} \rangle + \cdots + \lambda_n \langle \mathfrak{f}_n, \mathbf{u} \rangle = 0,$$

para todo  $\mathbf{u} \in X$ . Em particular, tomando  $\mathbf{u} = \mathbf{w}_i$ , para cada  $i = 1, \dots, n$ , e usando que  $\langle \mathfrak{f}_i, \mathbf{w}_j \rangle = 0$ , para  $j \neq i$ , e  $\langle \mathfrak{f}_i, \mathbf{w}_i \rangle = 1$ , obtemos

$$\lambda_1 \langle \mathfrak{f}_1, \mathbf{w}_i \rangle + \cdots + \lambda_n \langle \mathfrak{f}_n, \mathbf{w}_i \rangle = \lambda_i = 0.$$

Como  $i = 1, \dots, n$  é arbitrário, obtemos que  $\mathfrak{b}^*$  é LI. Portanto,  $\mathfrak{b}^*$  é uma base de  $X^*$ .  $\square$

A base  $\mathfrak{b}$  construída no [Teorema 3.1](#) leva o nome especial de base dual à  $\mathfrak{b}$ .

**DEFINIÇÃO 3.1** (Base dual). *Seja  $X$  um espaço vetorial de dimensão finita  $n \in \mathbb{N}$  e seja  $\mathfrak{b} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  uma base de  $X$ . Então a base  $\mathfrak{b}^* = \{\mathfrak{f}_1, \dots, \mathfrak{f}_n\}$  de  $X^*$  dada pelo [Teorema 3.1](#) é chamada de **base dual** à base  $\mathfrak{b}$ .*

**OBSERVAÇÃO 3.1.** Observe que, na passagem de [\(3.1\)](#) para [\(3.4\)](#) via [\(3.3\)](#), os coeficientes  $\lambda_i$  e os elementos da base  $\mathbf{w}_i$ , na representação de  $\mathbf{u}$ , trocaram de papel, virando, na representação de  $\mathfrak{f}$ , os elementos da base  $\mathfrak{f}_i$  e os coeficientes  $\alpha_i = \langle \mathfrak{f}, \mathbf{w}_i \rangle$ , respectivamente, em uma manifestação da *dualidade* entre  $X$  e  $X^*$ .

**COROLÁRIO 3.1.** *Seja  $X$  um espaço vetorial de dimensão finita. Então  $\dim(X^*) = \dim(X)$ .*

**DEMONSTRAÇÃO.** Caso  $X = \{\mathbf{0}\}$  seja o espaço trivial, então  $X^* = \{\mathfrak{o}\}$  e ambos têm dimensão zero. Caso  $X$  não seja trivial, então  $\dim(X) = n$  para algum  $n \in \mathbb{N}$ , de modo que o [Teorema 3.1](#) diz que  $X^*$  também têm dimensão  $n$ . Em qualquer caso, temos  $\dim(X) = \dim(X^*)$ .  $\square$

**OBSERVAÇÃO 3.2.** A base dual parece um tanto esdrúxula, à primeira vista, mas pode ser visualizada de maneira bem natural. Pense que os elementos da base dual são funcionais lineares e, portanto, podem ser visualizados pelas suas curvas de nível, que são hiperplanos no espaço primal. E são hiperplanos bem particulares, que estão associados diretamente ao sistema de coordenadas. Veja, por exemplo, a [Seção 3.2](#).

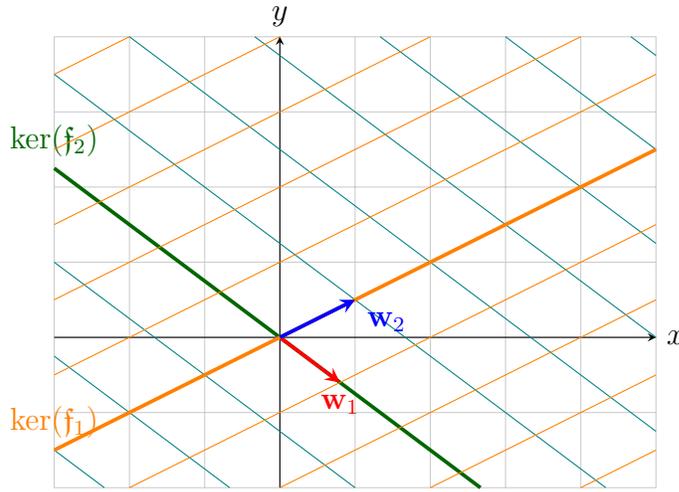


FIGURA 3.1. Ilustração do base dual  $\mathcal{b}^* = \{f_1, f_2\}$  através das curvas de nível de  $f_1$  e  $f_2$ , formando o sistema de coordenadas da base primal  $\mathcal{b} = \{w_1, w_2\}$ . Cada  $f_i$  da base dual  $\mathcal{b}^*$ , quando aplicado a um vetor  $\mathbf{u}$ , dá exatamente a  $i$ -ésima coordenada  $\langle f_i, \mathbf{u} \rangle$  da representação de  $\mathbf{u}$  na base  $\mathcal{b}$ . E cada funcional  $f$ , quando aplicado aos vetores  $w_i$  da base primal  $\mathcal{b}$ , dá a  $i$ -ésima coordenada  $\langle f, w_i \rangle$  de  $f$  na base dual. Observe que  $w_1$  está no núcleo de  $f_2$  e  $w_2$  está no núcleo de  $f_1$ .

**3.3. Representação de vetores e funcionais lineares.** Como corolário da demonstração do [Teorema 3.1](#), a base dual pode ser usada para dar uma expressão mais explícita para a representação de um vetor em uma base, feita na [Definição 2.17](#).

**TEOREMA 3.2** (Representação vetorial via base dual). *Seja  $X$  um espaço vetorial de dimensão finita  $n = \dim(X) \in \mathbb{N}$  e seja  $\mathcal{b} = \{w_1, \dots, w_n\}$  uma base de  $X$ . Seja  $\mathcal{b}^* = \{f_1, \dots, f_n\}$  a base dual correspondente, ou seja, tal que  $f_i(w_j) = 0$ , para  $i \neq j$ , e  $f_i(w_i) = 1$ . Então, dado  $\mathbf{u} \in X$ , temos*

$$\mathbf{u} = \langle f_1, \mathbf{u} \rangle w_1 + \dots + \langle f_n, \mathbf{u} \rangle w_n$$

e, em particular, a representação  $(\mathbf{u})_{\mathcal{b}}$  de um vetor  $\mathbf{u} \in X$  na base  $\mathcal{b}$  pode ser escrita como

$$(\mathbf{u})_{\mathcal{b}} = (\langle f_1, \mathbf{u} \rangle, \dots, \langle f_n, \mathbf{u} \rangle).$$

**DEMONSTRAÇÃO.** Segue da demonstração do [Teorema 3.1](#), mais precisamente das equações [\(3.1\)](#) e [\(3.2\)](#).  $\square$

Analogamente, a representação de um funcional linear na base dual é dada através da aplicação do funcional nos elementos da base primal.

**TEOREMA 3.3.** *Seja  $X$  um espaço vetorial de dimensão finita  $n = \dim(X) \in \mathbb{N}$ ,  $\mathfrak{b} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  uma base de  $X$  e  $\mathfrak{b}^* = \{\mathbf{f}_1, \dots, \mathbf{f}_n\}$  a base dual dada na **Definição 3.1**. Então os coeficientes de um funcional linear qualquer  $\mathbf{f} \in X^*$  nessa base dual  $\mathfrak{b}^*$  são dados por  $\langle \mathbf{f}, \mathbf{w}_j \rangle$ ,  $j = 1, \dots, n$ , de modo que*

$$\mathbf{f} = \langle \mathbf{f}, \mathbf{w}_1 \rangle \mathbf{f}_1 + \dots + \langle \mathbf{f}, \mathbf{w}_n \rangle \mathbf{f}_n.$$

**DEMONSTRAÇÃO.** Pelo **Teorema 3.2**, temos

$$\mathbf{u} = \langle \mathbf{f}_1, \mathbf{u} \rangle \mathbf{w}_1 + \dots + \langle \mathbf{f}_n, \mathbf{u} \rangle \mathbf{w}_n.$$

Assim, dado  $\mathbf{f} \in X^*$ ,

$$\langle \mathbf{f}, \mathbf{u} \rangle = \langle \mathbf{f}, \langle \mathbf{f}_1, \mathbf{u} \rangle \mathbf{w}_1 + \dots + \langle \mathbf{f}_n, \mathbf{u} \rangle \mathbf{w}_n \rangle.$$

Pela linearidade de  $\mathbf{f}$  e usando que os termos  $\langle \mathbf{f}_i, \mathbf{u} \rangle$  são escalares,

$$\langle \mathbf{f}, \mathbf{u} \rangle = \langle \mathbf{f}, \mathbf{w}_1 \rangle \langle \mathbf{f}_1, \mathbf{u} \rangle + \dots + \langle \mathbf{f}, \mathbf{w}_n \rangle \langle \mathbf{f}_n, \mathbf{u} \rangle.$$

Como isso vale para todo  $\mathbf{u} \in X$ , isso significa que, como funcionais lineares,

$$\mathbf{f} = \langle \mathbf{f}, \mathbf{w}_1 \rangle \mathbf{f}_1 + \dots + \langle \mathbf{f}, \mathbf{w}_n \rangle \mathbf{f}_n,$$

provando o resultado. □

Como consequência da representação de um funcional linear em relação à base dual, temos um isomorfismo entre o dual e  $\mathbb{K}^{1 \times n}$ .

**TEOREMA 3.4** (Representação do dual em  $\mathbb{K}^{1 \times n}$ ). *Seja  $X$  um espaço vetorial de dimensão finita  $n = \dim(X) \in \mathbb{N}$ ,  $\mathfrak{b} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  uma base de  $X$  e  $\mathfrak{b}^* = \{\mathbf{f}_1, \dots, \mathbf{f}_n\}$  a base dual dada na **Definição 3.1**. Então  $X^*$  é isomorfo a  $\mathbb{K}^{1 \times n}$  com o isomorfismo  $J : X^* \rightarrow \mathbb{K}^{1 \times n}$  sendo dado por*

$$J(\mathbf{f}) = [\mathbf{f}]_{\mathfrak{b}^*} = [\langle \mathbf{f}, \mathbf{w}_1 \rangle \quad \dots \quad \langle \mathbf{f}, \mathbf{w}_n \rangle] \in \mathbb{K}^{1 \times n}.$$

*O vetor-linha  $[\mathbf{f}]_{\mathfrak{b}^*}$  é uma representação de  $\mathbf{f}$  em  $\mathbb{K}^n$ , em relação a base  $\mathfrak{b}^*$ .*

**DEMONSTRAÇÃO.** Primeiramente, observe que, para  $\mathbf{f}, \mathbf{g} \in X^*$  e  $\lambda \in \mathbb{K}$ , temos, pela definição de adição e multiplicação por escalar em  $X^*$ , que

$$\begin{aligned} J(\mathbf{f} + \lambda \mathbf{g}) &= (\langle \mathbf{f} + \lambda \mathbf{g}, \mathbf{w}_1 \rangle, \dots, \langle \mathbf{f} + \lambda \mathbf{g}, \mathbf{w}_n \rangle) \\ &= (\langle \mathbf{f}, \mathbf{w}_1 \rangle + \lambda \langle \mathbf{g}, \mathbf{w}_1 \rangle, \dots, \langle \mathbf{f}, \mathbf{w}_n \rangle + \lambda \langle \mathbf{g}, \mathbf{w}_n \rangle) \\ &= (\langle \mathbf{f}, \mathbf{w}_1 \rangle, \dots, \langle \mathbf{f}, \mathbf{w}_n \rangle) + \lambda (\langle \mathbf{g}, \mathbf{w}_1 \rangle, \dots, \langle \mathbf{g}, \mathbf{w}_n \rangle) \\ &= J(\mathbf{f}) + \lambda J(\mathbf{g}). \end{aligned}$$

Portanto,  $J$  preserva as operações de adição e multiplicação por escalar. Falta mostrar que é uma bijeção. Primeiramente, segue da representação dada pelo [Teorema 3.4](#) que, se  $J(\mathbf{f}) = J(\mathbf{g})$  para  $\mathbf{f}, \mathbf{g} \in X^*$ , então

$$\langle \mathbf{f}, \mathbf{u} \rangle = J(\mathbf{f})(\mathbf{u})_{\mathcal{B}} = J(\mathbf{g})(\mathbf{u})_{\mathcal{B}} = \langle \mathbf{g}, \mathbf{u} \rangle,$$

para todo  $\mathbf{u} \in X$ , mostrando que  $\mathbf{f} = \mathbf{g}$ , ou seja,  $J$  é injetivo. Agora, dado  $\mathbf{f} = [f_1 \ \cdots \ f_n] \in \mathbb{K}^{1 \times n}$ , podemos definir  $\mathbf{f} \in X^*$  por

$$\langle \mathbf{f}, \mathbf{u} \rangle = \mathbf{f}(\mathbf{u})_{\mathcal{B}},$$

para todo  $\mathbf{u} \in \mathbb{K}^n$ .  $\mathbf{f}$  é, de fato, linear, pois o produto matriz-vetor é linear,

$$\langle \mathbf{f}, \mathbf{u} + \lambda \mathbf{v} \rangle = \mathbf{f}(\mathbf{u} + \lambda \mathbf{v})_{\mathcal{B}} = \mathbf{f}(\mathbf{u})_{\mathcal{B}} + \lambda \mathbf{f}(\mathbf{v})_{\mathcal{B}} = \langle \mathbf{f}, \mathbf{u} \rangle + \lambda \langle \mathbf{f}, \mathbf{v} \rangle,$$

Observe, ainda, que

$$\langle \mathbf{f}, \mathbf{w}_j \rangle = \mathbf{f}(\mathbf{w}_j)_{\mathcal{B}} = f_j,$$

de modo que

$$J(\mathbf{f}) = (\langle \mathbf{f}, \mathbf{w}_1 \rangle, \dots, \langle \mathbf{f}, \mathbf{w}_n \rangle) = (f_1, \dots, f_n) = \mathbf{f},$$

ou seja,  $J$  é sobrejetivo. Portanto,  $J$  é uma bijeção que preserva as operações de adição e multiplicação por escalar, o que faz de  $J$  um isomorfismo entre  $X^*$  e  $\mathbb{K}^{1 \times n}$ . Em particular, eles têm a mesma dimensão, ou seja,  $\dim(X^*) = n = \dim(\mathbb{K}^{1 \times n})$ .  $\square$

**OBSERVAÇÃO 3.3.** A [Observação 2.7](#) comenta que, em Geometria Diferencial, assim como em suas aplicações em Física Matemática, os vetores são chamados de *contra-variantes*. Por sua vez, os funcionais lineares  $\mathbf{f}$ , ou, melhor, os *co-vetores*  $[\mathbf{f}]_{\mathcal{B}^*}$ , que representam  $\mathbf{f}$  em uma base dual  $\mathcal{B}^*$ , são ditos *co-variantes*. Vejamos. Seja  $\mathcal{B} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  uma base de um espaço vetorial  $X$  de dimensão  $n \in \mathbb{N}$  e  $\mathcal{B}^* = \{f_1, \dots, f_n\}$  a base dual. Um dado  $\mathbf{f}$  pode ser expresso por

$$\mathbf{f} = y_1 f_1 + \cdots + y_n f_n,$$

onde  $y_i = \langle \mathbf{f}, \mathbf{w}_i \rangle$ . A mudarmos a base do espaço primal para  $\tilde{\mathcal{B}} = \{\tilde{\mathbf{w}}_1, \dots, \tilde{\mathbf{w}}_n\}$ , onde  $\tilde{\mathbf{w}}_i = \lambda \mathbf{w}_i$ , para algum  $\lambda > 0$ , a base dual é alterada para  $\tilde{\mathcal{B}}^* = \{\tilde{f}_1, \dots, \tilde{f}_n\}$  com

$$\tilde{f}_i = \frac{f_i}{\lambda},$$

para compensar o fato de que precisamos ter

$$\langle \tilde{f}_i, \tilde{\mathbf{w}}_i \rangle = \left\langle \frac{f_i}{\lambda}, \lambda \mathbf{w}_i \right\rangle = \langle f_i, \mathbf{w}_i \rangle = 1.$$

Desse modo,

$$\mathbf{f} = \lambda y_1 \frac{f_1}{\lambda} + \cdots + \lambda y_n \frac{f_n}{\lambda} = \lambda y_1 \tilde{f}_1 + \cdots + \lambda y_n \tilde{f}_n.$$

Ou seja, enquanto a base primal mudou de acordo com

$$\mathbf{w}_i \mapsto \lambda \mathbf{w}_i,$$

os coeficientes dos co-vetores mudaram da mesma forma,

$$y_i \mapsto \lambda y_i,$$

contrastando com a forma vista no **Observação 2.7** de que os coeficientes dos vetores mudaram de forma inversa,

$$x_i \mapsto \frac{x_i}{\lambda}.$$

Por isso os vetores são chamados de contra-variantes e os co-vetores de co-variantes.

Também podemos representar um elemento do dual em  $\mathbb{K}^n$ , visto que  $\mathbb{K}^{1 \times n}$  e  $\mathbb{K}^n$  são o mesmo espaço do ponto de vista vetorial. A diferença só aparece na multiplicação matricial.

**TEOREMA 3.5** (Representação do dual em  $\mathbb{K}^n$ ). *Seja  $X$  um espaço vetorial de dimensão finita  $n = \dim(X) \in \mathbb{N}$ ,  $\mathfrak{b} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  uma base de  $X$  e  $\mathfrak{b}^* = \{\mathbf{f}_1, \dots, \mathbf{f}_n\}$  a base dual dada na **Definição 3.1**. Então  $X^*$  é isomorfo a  $\mathbb{K}^n$  com o isomorfismo  $J: X^* \rightarrow \mathbb{K}^{1 \times n}$  sendo dado por*

$$J(\mathbf{f}) = [\mathbf{f}]_{\mathfrak{b}^*} = (\mathbf{f})_{\mathfrak{b}^*} = (\langle \mathbf{f}, \mathbf{w}_1 \rangle, \dots, \langle \mathbf{f}, \mathbf{w}_n \rangle) \in \mathbb{K}^n.$$

O vetor  $(\mathbf{f})_{\mathfrak{b}^*}$  é uma representação de  $\mathbf{f}$  em  $\mathbb{K}^n$ , em relação a base  $\mathfrak{b}^*$ .

**DEMONSTRAÇÃO.** Idem. □

**EXEMPLO 3.1.** Seja  $X = \mathbb{R}^2$  e considere a base canônica  $\mathbf{e} = \{\mathbf{e}_1, \mathbf{e}_2\}$ ,  $\mathbf{e}_1 = (1, 0)$  e  $\mathbf{e}_2 = (0, 1)$ . A base dual  $\mathfrak{b}^* = \{\mathbf{f}_1, \mathbf{f}_2\}$  é dada por  $\mathbf{f}_1(x, y) = x$ ,  $\mathbf{f}_2(x, y) = y$ , visto que

$$\mathbf{f}_1(1, 0) = 1, \mathbf{f}_1(0, 1) = 0, \mathbf{f}_2(1, 0) = 0, \mathbf{f}_2(0, 1) = 1.$$

Outra forma de ver isso é pela expansão de  $\mathbf{u} = (x, y)$  qualquer,

$$\mathbf{u} = (x, y) = x\mathbf{e}_1 + y\mathbf{e}_2 = \mathbf{f}_1(\mathbf{u})\mathbf{e}_1 + \mathbf{f}_2(\mathbf{u})\mathbf{e}_2.$$

**EXEMPLO 3.2.** Seja, novamente,  $X = \mathbb{R}^2$ , mas considere, agora, a base  $\mathfrak{b} = \{(1, 0), (1, 1)\}$ . Pela expansão de um vetor  $\mathbf{u} = (x, y)$  qualquer, temos

$$\mathbf{u} = (x, y) = (x - y)(1, 0) + y(1, 1),$$

de onde vemos que

$$\mathbf{f}_1(x, y) = x - y, \quad \mathbf{f}_2(x, y) = y.$$

formam a base dual  $\mathfrak{b}^* = \{\mathbf{f}_1, \mathbf{f}_2\}$ . Podemos verificar que

$$\mathbf{f}_1(1, 0) = 1, \mathbf{f}_1(1, 1) = 0, \mathbf{f}_2(1, 0) = 0, \mathbf{f}_2(1, 1) = 1.$$

Nessa base dual, o funcional  $f(x, y) = 2x + 3y$ , por exemplo, se escreve como

$$f(x, y) = f(1, 0)f_1 + f(1, 1)f_2 = 2f_1 + 5f_2.$$

Ou seja,

$$[f]_{\mathcal{f}^*} = [2 \ 5].$$

EXEMPLO 3.3. No espaço  $X = \mathcal{P}_3$  dos polinômios de grau menor ou igual a três, podemos considerar a base  $\mathcal{b} = \{1, x, x^2, x^3\}$ . A base dual é

$$\mathcal{b}^* = \left\{ \delta_0, \delta_0 D, \frac{1}{2}\delta_0 D^2, \frac{1}{6}\delta_0 D^3 \right\},$$

onde  $\delta_0$  é a delta de Dirac, que é o funcional linear  $\langle \delta_0, p \rangle = p(0)$ , e  $D$  representa a derivada em relação a  $x$ , de modo que

$$\langle \delta_0 D, p \rangle = p'(0), \quad \langle \frac{1}{2}\delta_0 D^2, p \rangle = \frac{1}{2}p''(0), \quad \langle \frac{1}{6}\delta_0 D^3, p \rangle = \frac{1}{6}p'''(0).$$

Não é difícil verificar que essa base satisfaz as condições de base dual, dadas no [Teorema 3.1](#).

EXEMPLO 3.4. Considerando o espaço  $X = \mathcal{P}_3$  e usando a base dual dada construída no [Exemplo 3.3](#), vamos ver como escrever o funcional  $\langle f, p \rangle = p(2)$  nessa base. Os coeficientes são dados por

$$\langle f, 1 \rangle = 1, \quad \langle f, x \rangle = 2, \quad \langle f, x^2 \rangle = 2^2 = 4, \quad \langle f, x^3 \rangle = 2^3 = 8.$$

Desse modo,

$$f = \delta_0 + 2\delta_0 D + \frac{4}{2}\delta_0 D^2 + \frac{8}{6}\delta_0 D^3.$$

Assim, para qualquer  $p \in \mathcal{P}_3$ ,

$$p(2) = p(0) + 2p'(0) + 2p''(0) + \frac{4}{3}p'''(0).$$

Por exemplo, no aplicando isso ao polinômio  $p(x) = 1 + 2x + 3x^3$ , encontramos, de um jeito ou de outro,  $p(2) = 29$ .

**3.4. Realização de um dual.** O dual é, naturalmente, um espaço um tanto mais abstrato. Vimos como representá-lo como  $\mathbb{K}^{1 \times n}$  ou  $\mathbb{K}^n$ , no caso de dimensão finita. Mas como fica o produto de dualidade? Ele é substituído por um funcional bilinear. Fazer isso em conjunto com o produto de dualidade nos leva ao conceito de *realização do dual*. Com ele, é possível dar uma representação mais concreta ao dual e ao produto de dualidade, o que também nos permite operar com esses objetos, deduzir resultados e obter estimativas de forma mais natural. Veja [Seção 3.4](#).

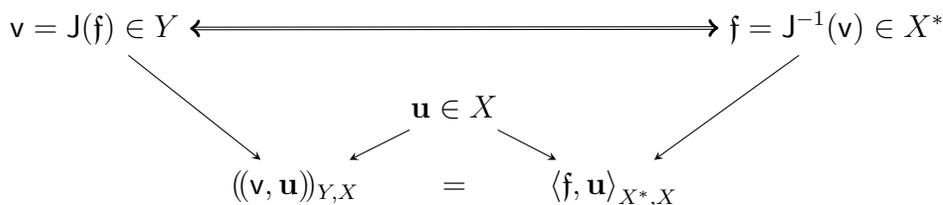


FIGURA 3.2. Realização do dual  $X^*$  de  $X$  via espaço vetorial  $Y$ , isomorfismo  $J : X^* \rightarrow Y$  e forma bilinear  $((\cdot, \cdot)) : Y \times X \rightarrow \mathbb{K}$ , com  $\langle \mathbf{f}, \mathbf{u} \rangle_{X^*,X} = ((J(\mathbf{f}), \mathbf{u}))_{Y,X}$ .

DEFINIÇÃO 3.2. *Seja  $X$  um espaço vetorial sobre um corpo  $\mathbb{K}$ . Uma **realização** do dual consiste em um espaço vetorial  $Y$  sobre o mesmo corpo  $\mathbb{K}$ , um isomorfismo  $J : X^* \rightarrow Y$  entre o dual  $X^*$  e  $Y$  e uma forma bilinear  $(\cdot, \cdot) : Y \times X \rightarrow \mathbb{K}$  tal que*

$$\langle \mathbf{f}, \mathbf{u} \rangle = (J(\mathbf{f}), \mathbf{u}), \quad (3.5)$$

para todo  $\mathbf{u} \in X$  e  $\mathbf{f} \in X^*$ .

EXEMPLO 3.5. Em alguns casos, para deixar claro os espaços envolvidos, denotamos a realização (3.5) por

$$\langle \mathbf{f}, \mathbf{u} \rangle_{X^*,X} = (J(\mathbf{f}), \mathbf{u})_{Y,X}.$$

Observe a ordem em que os espaços aparecem na notação, acompanhando a ordem dos argumentos do produto de dualidade e da forma bilinear.

O caso principal de realização do dual que exploraremos aqui é o do dual de  $\mathbb{K}^n$ , que iremos identificar, inicialmente, com vetores-linha  $\mathbb{K}^{1 \times n}$ , com a forma bilinear sendo o produto matriz-vetor. Outra representação que exploraremos posteriormente é a de representar o dual de  $\mathbb{K}^n$  pelo próprio  $\mathbb{K}^n$ , com a forma bilinear sendo o produto escalar. Isso pode ser feito para qualquer produto interno em  $\mathbb{K}^n$ , ou mais geralmente nos chamados *espaços de Hilbert*, que são espaços com produto interno que são completos em relação à norma gerada pelo produto interno.

TEOREMA 3.6. *Seja  $\mathbb{K}$  um corpo e considere o espaço vetorial  $X = \mathbb{K}^n$ . Então uma realização do dual  $X^*$  é o espaço  $Y = \mathbb{K}^{1 \times n}$ , com a forma bilinear sendo o produto matriz-vetor*

$$(v, u) = v u = y_1 x_1 + \cdots + y_n x_n,$$

onde

$$\mathbf{v} = [y_1 \ y_2 \ \cdots \ y_n] \in \mathbb{K}^{1 \times n}, \quad \mathbf{u} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{K}^n.$$

DEMONSTRAÇÃO. Considerando a base canônica  $\mathbf{e} = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$  de  $\mathbb{K}^n$ , com  $\mathbf{e}_1 = (1, 0, \dots, 0), \dots, \mathbf{e}_n = (0, \dots, 0, 1)$ , defimos o isomorfismo

$$\mathbf{J}(\mathbf{f}) = [\langle \mathbf{f}, \mathbf{e}_1 \rangle \ \langle \mathbf{f}, \mathbf{e}_2 \rangle \ \cdots \ \langle \mathbf{f}, \mathbf{e}_n \rangle].$$

Como  $\mathbf{u} = (x_1, \dots, x_n)$  pode ser escrito como

$$\mathbf{u} = x_1 \mathbf{e}_1 + \cdots + x_n \mathbf{e}_n,$$

temos, pela linearidade de  $\mathbf{f} \in X^*$ ,

$$\langle \mathbf{f}, \mathbf{u} \rangle = \langle \mathbf{f}, x_1 \mathbf{e}_1 + \cdots + x_n \mathbf{e}_n \rangle = x_1 \langle \mathbf{f}, \mathbf{e}_1 \rangle + \cdots + x_n \langle \mathbf{f}, \mathbf{e}_n \rangle = \mathbf{J}(\mathbf{f})\mathbf{u}.$$

□

OBSERVAÇÃO 3.4. Não deixemos que a aura da abstração encubra fatos simples. Por exemplo, se  $X = \mathbb{R}^2$  e  $\mathbf{f} \in X^*$ , então, escrevendo os elementos do espaço na forma  $\mathbf{u} = (x, y)$ , um funcional linear é uma função  $\mathbf{f}(x, y)$ . Tomando  $a = \mathbf{f}(1, 0)$ ,  $b = \mathbf{f}(0, 1)$  e escrevendo  $(x, y) = (x, 0) + (0, y) = x(1, 0) + y(0, 1)$ , obtemos, da linearidade de  $\mathbf{f}$ , que o funcional deve ser da forma

$$\mathbf{f}(x, y) = \mathbf{f}(x, 0) + \mathbf{f}(0, y) = x\mathbf{f}(1, 0) + y\mathbf{f}(0, 1) = ax + by = [a \ b] \begin{pmatrix} x \\ y \end{pmatrix}.$$

Analogamente (ou indutivamente), se  $X = \mathbb{R}^n$  e  $\mathbf{u} = (x_1, \dots, x_n)$ , então um funcional linear  $\mathbf{f} \in X^*$  deve ser da forma

$$\begin{aligned} \mathbf{f}(x_1, \dots, x_n) &= x_1 \mathbf{f}(1, 0, \dots, 0) + \cdots + x_n \mathbf{f}(0, \dots, 0, 1) = a_1 x_1 + \cdots + a_n x_n \\ &= [a_1 \ \cdots \ a_n] \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}. \end{aligned}$$

**3.5. Representação do espaço e do dual.** Também podemos representar qualquer espaço de dimensão finita por  $\mathbb{K}^n$ , de modo que a representação feita no [Teorema 3.6](#) pode ser estendida a um espaço de dimensão finita qualquer. Podemos representar tanto  $X$  como  $X^*$  através de suas isomorfias como  $\mathbb{K}^n$  e  $\mathbb{K}^{1 \times n}$  ou mesmo ambos com  $\mathbb{K}^n$ . Sendo assim, a aplicação de um funcional  $\mathbf{f} \in X^*$  em um vetor  $\mathbf{u} \in X$  se reflete em uma operação entre os seus representantes. Veja [Seção 3.5](#)

TEOREMA 3.7. Dada uma base  $\mathcal{b} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  de um espaço vetorial  $X$  de dimensão finita  $n \in \mathbb{N}$  sobre um corpo  $\mathbb{K}$  e  $\mathcal{b}^* = \{\mathbf{f}_1, \dots, \mathbf{f}_n\}$  a base dual, a aplicação de um funcional linear  $\mathbf{f} \in X^*$  a um vetor  $\mathbf{u} \in X$  pode ser expressa via produto matriz-vetor entre as suas representações em  $\mathbb{K}^n$  e  $\mathbb{K}^{1 \times n}$ , respectivamente, i.e.

$$\langle \mathbf{f}, \mathbf{u} \rangle = [\mathbf{f}]_{\mathcal{b}^*}(\mathbf{u})_{\mathcal{b}} = f_1 u_1 + \dots + f_n u_n,$$

onde

$$[\mathbf{f}]_{\mathcal{b}^*} = [f_1 \ \dots \ f_n] = [\langle \mathbf{f}, \mathbf{w}_1 \rangle \ \dots \ \langle \mathbf{f}, \mathbf{u} \rangle] \in \mathbb{K}^{1 \times n}$$

e

$$(\mathbf{u})_{\mathcal{b}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \langle \mathbf{f}_1, \mathbf{u} \rangle \\ \vdots \\ \langle \mathbf{f}_n, \mathbf{u} \rangle \end{pmatrix} \in \mathbb{K}^n.$$

DEMONSTRAÇÃO. Basta observar que, pela linearidade de  $\mathbf{f} \in Y^*$ ,

$$\langle \mathbf{f}, \mathbf{u} \rangle = \langle \mathbf{f}, x_1 \mathbf{w}_1 + \dots + x_n \mathbf{w}_n \rangle = x_1 \langle \mathbf{f}, \mathbf{w}_1 \rangle + \dots + x_n \langle \mathbf{f}, \mathbf{w}_n \rangle = [\mathbf{f}]_{\mathcal{b}^*} \cdot (\mathbf{u})_{\mathcal{b}}.$$

□

No caso em que ambos  $X$  e  $X^*$  são identificados com  $\mathbb{K}^n$ , a forma bilinear é, de fato, o produto escalar. Isso antecipa um resultado mais geral de representação de funcionais lineares em espaços de Hilbert (que são espaços vetoriais com produto interno e completos em relação à topologia da norma associada ao produto interno), conhecido como o *Teorema da Representação de Riesz*. Vejamos, a seguir, esse resultado no nosso caso particular de dimensão finita.

TEOREMA 3.8. Dada uma base  $\mathcal{b} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  de um espaço vetorial  $X$  de dimensão finita  $n \in \mathbb{N}$  sobre um corpo  $\mathbb{K}$  e  $\mathcal{b}^* = \{\mathbf{f}_1, \dots, \mathbf{f}_n\}$  a base dual, a aplicação de um funcional linear  $\mathbf{f} \in X^*$  a um vetor  $\mathbf{u} \in X$  pode ser expressa via produto escalar entre as suas representações em cada base, i.e.

$$\langle \mathbf{f}, \mathbf{u} \rangle = (\mathbf{f})_{\mathcal{b}^*} \cdot (\mathbf{u})_{\mathcal{b}} = \langle \mathbf{f}, \mathbf{w}_1 \rangle u_1 + \dots + \mathbf{f}(\mathbf{w}_n) u_n,$$

onde

$$(\mathbf{f})_{\mathcal{b}^*} = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} = [\langle \mathbf{f}, \mathbf{w}_1 \rangle \ \dots \ \langle \mathbf{f}, \mathbf{u} \rangle] \in \mathbb{K}^n$$

e

$$(\mathbf{u})_{\mathcal{b}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \langle \mathbf{f}_1, \mathbf{u} \rangle \\ \vdots \\ \langle \mathbf{f}_n, \mathbf{u} \rangle \end{pmatrix} \in \mathbb{K}^n.$$

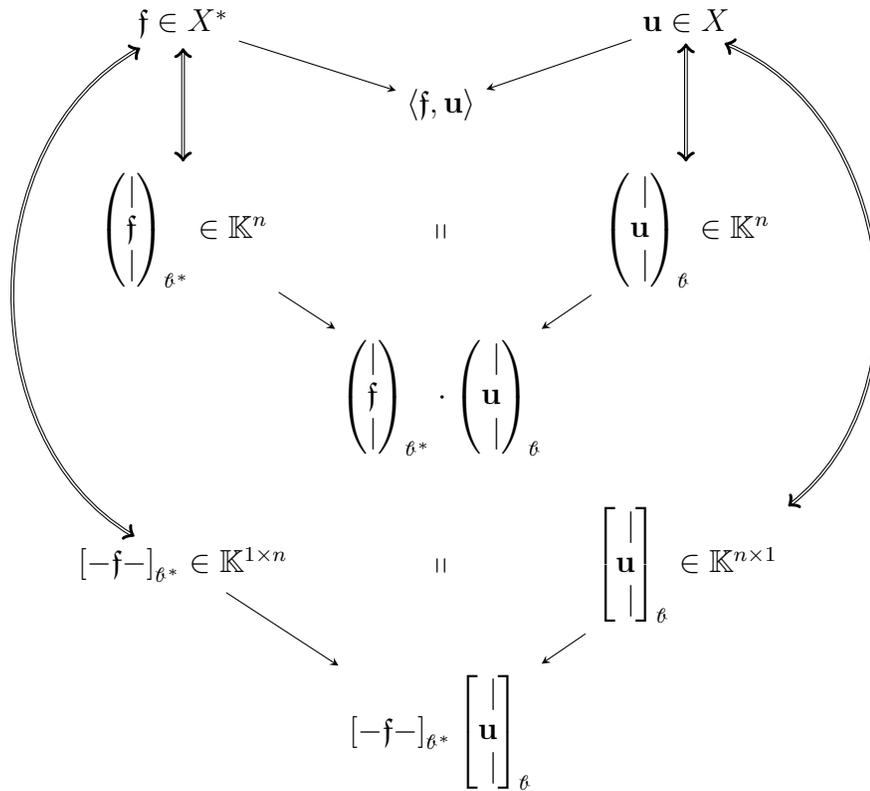


FIGURA 3.3. Diferentes representações para  $X$  e  $X^*$  e para o produto de dualidade  $\langle \mathbf{f}, \mathbf{u} \rangle$ . Dada uma base  $\mathcal{b}$  para  $X$ , com base dual associada  $\mathcal{b}^*$  para  $X^*$ , cada elemento  $\mathbf{u} \in X$  pode ser representado pelo vetor  $(\mathbf{u})_{\mathcal{b}} \in \mathbb{K}^n$  dos coeficientes de  $\mathbf{u}$  na base  $\mathcal{b}$  ou por um vetor-coluna  $[\mathbf{u}]_{\mathcal{b}} \in \mathbb{K}^{n \times 1}$ , enquanto que cada  $\mathbf{f} \in X^*$  pode ser representado pelo vetor  $(\mathbf{f})_{\mathcal{b}^*}$  dos coeficientes de  $\mathbf{f}$  na base dual  $\mathcal{b}^*$  ou pelo vetor-linha  $[\mathbf{f}]_{\mathcal{b}^*}$ . Em  $\mathbb{K}^n \times \mathbb{K}^n$ , o produto de dualidade toma a forma de um produto escalar  $(\mathbf{f})_{\mathcal{b}^*} \cdot (\mathbf{u})_{\mathcal{b}}$ , enquanto que em  $\mathbb{K}^{n \times 1} \times \mathbb{K}^{1 \times n}$ , o produto de dualidade toma a forma de uma multiplicação matricial  $[\mathbf{f}]_{\mathcal{b}^*} [\mathbf{u}]_{\mathcal{b}}$ .

DEMONSTRAÇÃO. Análoga a do Teorema 3.7.

□

#### 4. O bidual

Como  $X^*$  é um espaço vetorial, podemos considerar, também, o seu dual, que é chamado de bidual de  $X$ .

##### 4.1. Definição.

DEFINIÇÃO 4.1 (Bidual). *Dado um espaço vetorial  $X$ , o seu **bidual**, denotado por  $X^{**}$ , é o dual do seu dual, ou seja  $X^{**} = (X^*)^*$ .*

EXEMPLO 4.1. Seja  $X = \mathbb{R}^2$  e considere um funcional linear qualquer

$$f(\mathbf{u}) = f(x, y) = ax + by,$$

agindo em  $\mathbf{u} = (x, y) \in \mathbb{R}^2$ . Fixando um vetor  $\mathbf{u}_0 = (x_0, y_0)$  arbitrário, a função

$$f \mapsto f(\mathbf{u}_0) = ax_0 + by_0$$

é um funcional linear, que leva um elemento  $f = f(x, y) = ax + by$  de  $X^*$  no escalar  $f(\mathbf{u}_0) = ax_0 + by_0$ . Ele é, portanto, um elemento do dual do dual de  $X$ , i.e. um elemento de  $X^{**}$ .

##### 4.2. Morfismo canônico.

OBSERVAÇÃO 4.1. O funcional do [Exemplo 4.1](#) não é um mero exemplo, mas um caso típico, que nos leva a uma identificação entre cada elemento de um espaço em um elemento do bidual. Isso define uma transformação linear que é chamada de *morfismo canônico* ou *morfismo natural*. No caso de dimensão finita, veremos, em seguida, que esse morfismo é um isomorfismo.

DEFINIÇÃO 4.2. *Seja  $X$  um espaço vetorial. O **morfismo natural** ou **canônico** é a transformação linear  $J : X \rightarrow X^{**}$  definida através do produto de dualidade, levando um vetor  $\mathbf{u} \in X$  em um elemento  $J(\mathbf{u}) \in X^{**}$  que é o funcional linear em  $X^*$  dado por*

$$\langle J(\mathbf{u}), f \rangle_{X^{**}, X^*} = \langle f, \mathbf{u} \rangle_{V^*, V}.$$

OBSERVAÇÃO 4.2. É imediato verificar que o morfismo natural  $J$  dado na [Definição 4.2](#) é linear. De fato,

$$\begin{aligned} \langle J(\mathbf{u} + \lambda\mathbf{v}), f \rangle &= \langle f, \mathbf{u} + \lambda\mathbf{v} \rangle = \langle f, \mathbf{u} \rangle + \lambda\langle f, \mathbf{v} \rangle \\ &= \langle J(\mathbf{u}), f \rangle + \lambda\langle J(\mathbf{v}), f \rangle = \langle J(\mathbf{u}) + \lambda J(\mathbf{v}), f \rangle, \end{aligned}$$

para todo  $f \in X^*$ , ou seja,

$$J(\mathbf{u} + \lambda\mathbf{v}) = J(\mathbf{u}) + \lambda J(\mathbf{v}).$$

**4.3. Isomorfismo natural com o bidual em dimensão finita.** Como mencionado acima, no caso de dimensão finita, o morfismo natural é um isomorfismo entre o espaço e o seu bidual.

**TEOREMA 4.1.** *Seja  $X$  um espaço vetorial de dimensão finita. O bidual  $X^{**}$  é isomorfo a  $X$  através do morfismo natural, ou canônico, dado na [Definição 4.2](#).*

**DEMONSTRAÇÃO.** Se  $X = \{\mathbf{0}\}$  é o espaço trivial, é fácil ver que tanto  $X^*$  quanto  $X^{**}$  são triviais, de modo que o isomorfismo é imediato. Considere, agora,  $X$  não trivial. Vamos mostrar que  $J$  é uma bijeção.

Se  $J(\mathbf{u}) = J(\mathbf{v})$ , então  $\langle J(\mathbf{u}), \mathbf{f} \rangle = \langle J(\mathbf{v}), \mathbf{f} \rangle$  para todo  $\mathbf{f} \in X$ , ou seja

$$\langle \mathbf{f}, \mathbf{u} \rangle = \langle J(\mathbf{u}), \mathbf{f} \rangle = \langle J(\mathbf{v}), \mathbf{f} \rangle = \langle \mathbf{f}, \mathbf{v} \rangle,$$

para todo  $\mathbf{f} \in X$ . Segue, então, do [Corolário 2.4](#), que  $\mathbf{u} = \mathbf{v}$ , mostrando que  $J$  é injetivo.

Para ver que  $J : X \rightarrow X^{**}$  é sobrejetivo, vamos primeiro mostrar que, dada uma base  $\mathcal{b} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  de  $X$ , segue que  $\{J(\mathbf{w}_1), \dots, J(\mathbf{w}_n)\}$  é uma base para  $X^{**}$ .

Considere a base dual  $\mathcal{b}^* = \{\mathbf{f}_1, \dots, \mathbf{f}_n\}$  de  $X^*$ , onde  $\langle \mathbf{f}_i, \mathbf{w}_j \rangle = \delta_{ij}$  é o Delta de Kronecker. Observe que cada  $J(\mathbf{w}_i)$  é tal que

$$\langle J(\mathbf{w}_i), \mathbf{f}_j \rangle = \langle \mathbf{f}_j, \mathbf{w}_i \rangle = \delta_{ij}.$$

Portanto, segue do [Teorema 3.1](#) que  $\mathcal{b}^{**} = \{J(\mathbf{w}_1), \dots, J(\mathbf{w}_n)\}$  é base do dual de  $X^*$ , ou seja, é base do bidual  $X^{**}$ .

Em particular, dada  $\mathbf{g} \in X^{**}$ , podemos escrever

$$\mathbf{g} = \gamma_1 J(\mathbf{w}_1) + \dots + \gamma_n J(\mathbf{w}_n) = J(\gamma_1 \mathbf{w}_1 + \dots + \gamma_n \mathbf{w}_n) = J(\mathbf{u}),$$

para  $\mathbf{u} = \gamma_1 \mathbf{w}_1 + \dots + \gamma_n \mathbf{w}_n$ , ou seja,  $J : X \rightarrow X^{**}$  é sobrejetivo.

Portanto,  $J$  é uma bijeção que preserva a estrutura de espaço vetorial, ou seja, é um isomorfismo entre  $X$  e  $X^{**}$ , como queríamos mostrar.  $\square$

**OBSERVAÇÃO 4.3.** Em dimensão finita, é fácil identificar o dual  $X^*$  com o próprio  $X$  e com o seu bidual  $X^{**}$ , além do dual algébrico coincidir com o dual topológico. Em dimensão infinita, as coisas não são tão simples. Essa identificação entre o espaço e o espaço dual se estende para o dual topológico quando o espaço é um espaço de Hilbert. Mas para espaços de Banach ou outros espaços vetoriais topológicos, o dual topológico  $X'$  pode ser de uma natureza bastante diferente de  $X$ . A identificação entre o espaço  $X$  e o seu bidual topológico  $X''$ , por sua vez, já vale para uma gama mais ampla de espaços, chamados de espaços *reflexivos*, para os quais o morfismo canônico  $J : X \rightarrow X^{**}$  tem um papel crucial. Esses espaços são ditos reflexivos justamente quando esse morfismo é um isomorfismo.

OBSERVAÇÃO 4.4. Apenas para saciar, ou instigar, um pouco a curiosidade, tomando  $1 < p < \infty$ , o dual topológico de  $\ell^p$  pode ser representado pelo espaço  $\ell^q$ , onde  $q$  é o “dual” de  $p$ , definido por  $1/p + 1/q = 1$ , com o produto de dualidade dado por

$$\langle \mathbf{v}, \mathbf{u} \rangle_{\ell^q, \ell^p} = \sum_j v_j u_j,$$

onde  $\mathbf{u} = (u_1, u_2, \dots) \in \ell^p$  e  $\mathbf{v} = (v_1, v_2, \dots) \in \ell^q$ . Observe, por conseguinte, que o dual topológico de  $\ell^q$  é, de volta, o próprio  $\ell^p$ . Em particular, esses são espaços *reflexivos*, onde o bidual é isomorfo ao espaço. No caso  $p = 1$ , o dual topológico de  $\ell^1$  é  $\ell^\infty$ . Porém, o dual de  $\ell^\infty$ , que é o bidual de  $\ell^1$ , não é o  $\ell^1$ , sendo um espaço um tanto mais complicado, o que faz de  $\ell^1$  um exemplo de espaço não-reflexivo (assim como  $\ell^\infty$ ).

## 5. Anulador de um subespaço

**5.1. O subespaço anulador.** Vimos que o núcleo de um funcional linear não nulo é um subespaço de co-dimensão um. Se  $\mathbf{f} \in X^*$  é não nulo e  $S = \ker(\mathbf{f})$ , podemos dizer que  $\mathbf{f}$  “anula” os elementos de  $S$ . Mais geralmente, temos a seguinte definição.

DEFINIÇÃO 5.1 (Anulador). *Seja  $S$  um subespaço de um espaço vetorial  $X$ . O **anulador** de  $S$  é o conjunto de todos os funcionais lineares que se anulam em  $S$  e é denotado por  $S^0$  (ou  $S^\perp$ ), i.e.*

$$S^0 = \{\mathbf{f} \in X^*; \langle \mathbf{f}, \mathbf{v} \rangle = 0, \forall \mathbf{v} \in S\}.$$

A [Seção 5.1](#) ilustra o conjunto de anuladores de um hiperplano  $S$ . Escolhendo qualquer  $\mathbf{w} \in X \setminus S$ , podemos definir um funcional  $\mathbf{f} \in X^*$  que anula  $S$  e tal que  $\langle \mathbf{f}, \mathbf{w} \rangle = 1$ . Variando  $\mathbf{w}$  (ou variando o valor  $\langle \mathbf{f}, \mathbf{w} \rangle = \lambda$ ), obtemos uma família unidimensional de funcionais, cujos “gráficos” são “hiperplanos” no espaço estendido  $X \times \mathbb{K}$  (com a inclinação desses hiperplanos diretamente relacionada a  $\lambda$ ). O funcional nulo  $\mathbf{o} \in X^*$  também anula  $S$  e o seu gráfico é o próprio  $X \times \{0\}$ .

OBSERVAÇÃO 5.1. Não há produto interno para justificar a notação de  $S^0$  com sobrescrito  $\perp$ , até mesmo porque  $\mathbf{f}$  e  $\mathbf{u}$  “moram” em espaços distintos, mas a notação é baseada no produto de dualidade, em que  $\langle \mathbf{f}, \mathbf{u} \rangle = 0$  para  $\mathbf{u} \in S$  e  $\mathbf{f} \in S^0$ , que se assemelha ao de um produto interno. É verdade, no entanto, que quando  $X$  possui um produto interno  $(\cdot, \cdot)$  e é de dimensão finita ou, mais geralmente, é completo na norma associada ao produto interno, então todo  $\mathbf{f} \in X'$  (o dual topológico) possui um “representante”  $\mathbf{v}$  em  $X$  tal que  $\langle \mathbf{f}, \mathbf{u} \rangle = (\mathbf{v}, \mathbf{u})$ , para todo  $u \in X$ . O conjunto desses representantes é, de fato, ortogonal a  $S$ , segundo o produto interno  $(\cdot, \cdot)$ , e é comumente denotado por  $S^\perp$ , que é um subespaço de  $X$ , enquanto que  $S^0$  é um

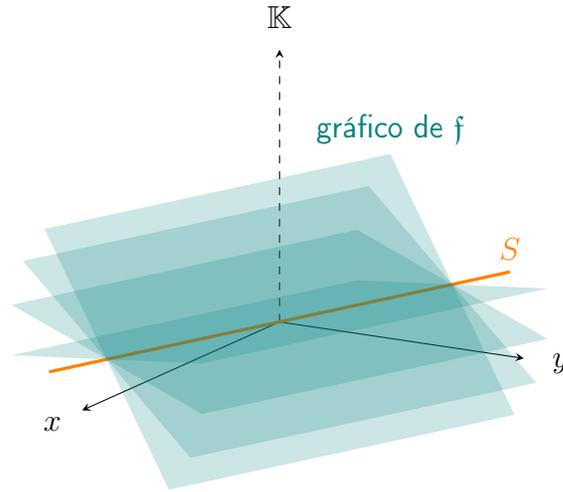


FIGURA 5.1. Ilustração da família  $S^0 = \{f \in X^*; S \subset \ker(f)\}$  de anuladores de um hiperplano  $S$  (unidimensional, nesse caso), em um espaço vetorial  $X$  representado pelo plano  $xy$ . Os anuladores estão representados através de seus gráficos, que no caso são planos no espaço  $xyz$ , onde  $z = f(\mathbf{u})$ ,  $\mathbf{u} = (x, y)$ . Mais geralmente, os anuladores são todos os hiperplanos, no espaço estendido  $X \times \mathbb{K}$ , que contém  $S \times \{0\} \subset X \times \mathbb{K}$ .

subespaço de  $X'$  isomorfo a  $S^\perp$ . Por conta disso, alguns autores usam  $S^\perp$  para denotar o que chamamos de  $S^0$ .

PROPOSIÇÃO 5.1. *O anulador  $S$  definido na Definição 5.1 é um subespaço vetorial de  $X^*$ .*

DEMONSTRAÇÃO. Imediata. Verifique. □

## 5.2. Dimensão do anulador.

TEOREMA 5.1. *Seja  $X$  um subespaço vetorial de dimensão finita e seja  $S \subset X$  um subespaço de  $X$ . Então*

$$\dim(X) = \dim(S) + \dim(S^0).$$

DEMONSTRAÇÃO. Se  $S = \{\mathbf{0}\}$  é o subespaço trivial, então qualquer funcional linear se anula em  $S$ , ou seja,  $S^0 = X^*$ . Nesse caso,  $\dim(S) = 0$  e  $\dim(S^0) = \dim(X)$ , provando o resultado nesse caso.

Suponha, agora, que  $S$  tenha dimensão  $k \in \mathbb{N}$ , com, naturalmente,  $k \leq n = \dim(X)$ . Se  $k = n$ , então  $S = X$  e  $S^0 = \{0\}$ , nos dando, novamente,  $\dim(X) = \dim(S) = \dim(S) + \dim(S^0)$ .

Se  $k < n$ , então  $S$  possui um subespaço complementar  $S'$  de dimensão  $n - k$ . Seja  $\{\mathbf{w}_1, \dots, \mathbf{w}_k\}$  uma base de  $S$  e seja  $\{\mathbf{w}_{k+1}, \dots, \mathbf{w}_n\}$  uma base de  $S'$ . Pelo **Lema 3.1** e pelo **Teorema 3.1**, existem funcionais  $f_1, \dots, f_n$  tais que  $f_i(\mathbf{w}_j) = \delta_{ij}$ , de modo que  $\{f_1, \dots, f_n\}$  forma uma base de  $X^*$ . Nenhum  $f_1, \dots, f_k$  anula  $S$ , enquanto que todos os  $f_{k+1}, \dots, f_n$  anulam  $S$ . Dessa forma,

$$S^0 = \text{span}\{f_{k+1}, \dots, f_n\},$$

de modo que  $\dim(S_0) = n - k$ . Com isso

$$\dim(X) = n = k + n - k = \dim(S) + \dim(S^0),$$

completando a demonstração. □

## 6. Aplicações

### 6.1. Quadratura de polinômios.

**TEOREMA 6.1** (Quadratura de polinômios). *Considere o espaço vetorial  $\mathcal{P}_n(\mathbb{R})$  dos polinômios de grau menor ou igual a  $n \in \mathbb{N}$ . Seja  $I \subset \mathbb{R}$  um intervalo finito e considere  $n + 1$  pontos distintos  $x_0, \dots, x_n$  em  $I$ . Então existem números reais  $m_0, \dots, m_n$  tais que*

$$\int_I p(x) \, dx = m_0 p(x_0) + \dots + m_n p(x_n), \quad (6.1)$$

para todo  $p \in \mathcal{P}_n(\mathbb{R})$ .

**DEMONSTRAÇÃO.** Observe que

$$p \mapsto f(p) = \int_I p(x) \, dx$$

define um funcional linear em  $\mathcal{P}_n(\mathbb{R})$ , ou seja,  $f \in \mathcal{P}_n(\mathbb{R})^*$  pertence ao dual algébrico desse espaço. A ideia é mostrar que os funcionais  $p \mapsto p(x_j)$  formam uma base para esse dual, de modo que  $f$  se escreve como combinação linear desses funcionais.

Defina, então, para cada  $j = 0, \dots, n$ , o funcional linear  $f_j \in \mathcal{P}_n(\mathbb{R})^*$  dado por

$$f_j(p) = p(x_j),$$

para  $p \in \mathcal{P}_n(\mathbb{R})$ . Vamos mostrar que  $\mathcal{B} = \{f_0, \dots, f_n\}$  é LI. Suponha que

$$\lambda_0 f_0 + \dots + \lambda_n f_n = 0.$$

Então

$$\lambda_0 f_0(p) + \cdots + \lambda_n f_n(p) = 0,$$

para todo  $p \in \mathcal{P}_n(\mathbb{R})$ . Pela definição de  $f_j$ , isso significa que

$$\lambda_0 p(x_0) + \cdots + \lambda_n p(x_n) = 0,$$

para todo  $p \in \mathcal{P}_n(\mathbb{R})$ . Para cada  $j = 0, \dots, n$ , considere o polinômio

$$q_j(x) = \prod_{i \neq j} (x - x_i).$$

Observe que  $q_j$  é um polinômio de grau  $n$ , portanto pertence a  $\mathcal{P}_n(\mathbb{R})$ . Além disso,  $q_j(x_i) = 0$ , para  $i \neq j$  e  $q_j(x_j) \neq 0$ . Com isso,

$$0 = \lambda_0 q_j(x_0) + \cdots + \lambda_n q_j(x_n) = \lambda_j q_j(x_j),$$

o que nos dá

$$\lambda_j = 0.$$

Como isso vale para todo  $j = 0, \dots, n$ , vemos que  $\mathfrak{b}^*$  é LI. Como  $\mathcal{P}_n(\mathbb{R})^*$  tem dimensão  $n + 1$ , segue que  $\mathfrak{b}^*$  é uma base de  $\mathcal{P}_n(\mathbb{R})^*$ .

Agora, considere a quadratura

$$f(p) = \int_I p(x) \, dx.$$

Esse é um funcional linear em  $\mathcal{P}(\mathbb{R})^*$ . Portanto, ele se escreve de forma única como combinação linear dos elementos da base  $\mathfrak{b}^*$ . Sendo assim, existem escalares  $m_0, \dots, m_n$  tais que

$$\int_I p(x) \, dx = f(p) = m_0 f_0(p) + \cdots + m_n f_n(p) = m_0 p(x_0) + \cdots + m_n p(x_n),$$

completando a demonstração. □

**OBSERVAÇÃO 6.1.** Quem são os coeficientes  $m_j$  em (6.1)? Para ver isso, aplicamos a fórmula aos polinômios  $q_j$  definidos na demonstração do **Teorema 6.1**. Como  $q_j(x_i) = 0$  para  $j \neq i$ , temos

$$\int_I q_j(x) \, dx = m_j q_j(x_j),$$

ou seja,

$$m_j = \int_I \frac{q_j(x)}{q_j(x_j)} \, dx.$$

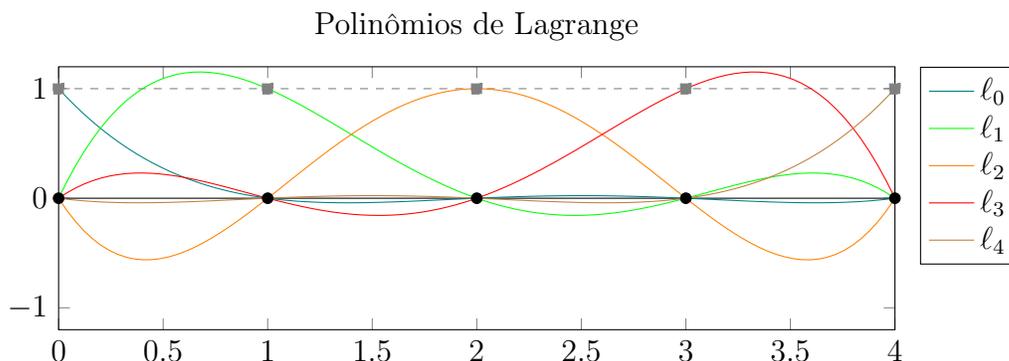


FIGURA 6.1. Polinômios de Lagrange  $\ell_j = \prod_{i \neq j} (x - x_i) / \prod_{i \neq j} (x_j - x_i)$ , para  $j = 0, \dots, 4$ , no intervalo real  $[0, 4]$ . Cada  $\ell_j$  se anula em  $x = x_i$ , para  $i \neq j$ , e vale 1 em  $x = x_j$ .

Os integrandos são os polinômios conhecidos como polinômios de Lagrange,

$$\ell_j(x) = \frac{\prod_{i \neq j} (x - x_i)}{\prod_{i \neq j} (x_i - x_j)},$$

que têm, ainda, a propriedade  $\ell_j(x_j) = 1$ . Ou seja, são uma versão normalizada dos  $q_j$ . Cada coeficiente  $m_j$  é a quadratura de cada polinômio de Lagrange  $\ell_j$ , no intervalo  $I$ .

**6.2. Subdiferenciais de funções convexas.** Por vir...

**6.3. Classificação de dados via máquina de vetores de suporte.** Por vir...

## 7. Exercícios

### Exercícios

**7.1.** Faça os detalhes da demonstração da **Proposição 2.5**.

**7.2.** Seja  $X = \mathbb{R}^2$  e considere os conjuntos dados por  $A = \{(0, 0), (1, 0)\}$  e  $B = \{(0, 1), (0, 2)\}$ . Encontre um hiperplano  $H = \{f = \eta\}$ , com  $f \in X^*$  (pode pensar na forma  $f(\mathbf{u}) = ax + by$ , para  $\mathbf{u} = (x, y)$ ) que separe os conjuntos de tal forma que  $A \subset \{f \leq \eta - \varepsilon\}$ ,  $B \subset \{f \geq \eta + \varepsilon\}$ , para algum  $\varepsilon > 0$  “ótimo”, ou seja, tal que nenhuma das duas inclusões valha para  $\varepsilon' > \varepsilon$ . Esse exercício é inspirado nos modelos de *máquina de vetores de suporte* (*support vector machine* - SVM). (O hiperplano pode ser escolhido de forma única impondo as condições de  $a^2 + b^2 = 1$  e de  $\varepsilon$  ser o maior possível.)

- 7.3.** Com  $X$ ,  $A$  e  $B$  como no exercício anterior, encontre um hiperplano  $H$  tal que todos os pontos do hiperplano sejam equidistantes das centroides de  $A$  e de  $B$ . Esse exercício é inspirado no método de  $k$ -means, com  $k = 2$ .
- 7.4.** Considere a base canônica  $\mathcal{b} = \{1, x, x^2, x^3\}$  do espaço  $\mathcal{P}_3(\mathbb{R})$  dos polinômios reais de grau no máximo quatro. Encontre explicitamente os elementos da base dual  $\mathcal{b}^* = \{f_0, f_1, f_2, f_3\}$ , tais que  $\langle f_i, x^i \rangle = 1$  e  $\langle f_i, x^j \rangle = 0$ ,  $i, j = 0, \dots, 3$ ,  $j \neq i$ .
- 7.5.** Idem para a base de polinômios de Lagrange  $\mathcal{a} = \{\ell_0, \ell_1, \ell_2, \ell_3\}$ , em pontos distintos  $x_0 < x_1 < x_2 < x_3$ , onde  $\ell_i = \prod_{j \neq i} (x - x_j) / \prod_{j \neq i} (x_i - x_j)$ .
- 7.6.** Usando a base dual  $\mathcal{b}^*$  da base  $\mathcal{b} = \{1, x, x^2, x^3\}$  do espaço  $\mathcal{P}_3(\mathbb{R})$ , feita no **Exercício 7.4**, represente os seguintes funcionais lineares
- $\langle f, p \rangle = p(0)$ ;
  - $\langle f, p \rangle = p(1)$ ;
  - $\langle f, p \rangle = p'(2)$ ;
  - $\langle f, p \rangle = \int_0^1 p(s) ds$ .
- 7.7.** Represente os mesmos funcionais lineares do **Exercício 7.6**, agora na base dual da base de polinômios de Lagrange, obtida no **Exercício 7.5**, tomando os pontos  $x_0 = 0, x_1 = 1, x_2 = 2, x_3 = 3$ .
- 7.8.** Encontre os funcionais  $\langle f_1, (x, y) \rangle = f_1(x, y) = ax + by$  e  $\langle f_2, (x, y) \rangle = f_2(x, y) = cx + dy$  da base dual  $\mathcal{b}^* = \{f_1, f_2\}$  associada à base  $\mathcal{b} = \{(1, 1), (-1, 1)\}$  de  $\mathbb{R}^2$ .
- 7.9.** Encontre explicitamente os funcionais lineares  $\langle f_i, (x, y, z) \rangle = f_i(x, y, z) = a_i x + b_i y + c_i z$ ,  $a_i, b_i, c_i \in \mathbb{R}$ ,  $i = 1, 2, 3$ , e  $\langle f_2, (x, y) \rangle = cx + dy$  da base dual  $\mathcal{b}^* = \{f_1, f_2, f_3\}$  associada à base  $\mathcal{b} = \{(1, 0, 0), (2, 1, 0), (5, 1, 1)\}$  de  $\mathbb{R}^3$ .
- 7.10.** Os funcionais lineares  $f_1(x, y) = \langle f_1, (x, y) \rangle = 2x - 3y$  e  $f_2(x, y) = \langle f_2, (x, y) \rangle = 4x - y$  formam uma base  $\mathcal{b}^*$  do dual de  $X = \mathbb{R}^2$ . Encontre a base primal  $\mathcal{b}$  de  $X$ , associada a essa base dual.
- 7.11.** Seja  $X$  um espaço vetorial real com relação de ordem  $\leq$ , conforme definido no **Exercício 5.13** do **Capítulo 1**. Essa relação de ordem induz uma relação no dual  $X^*$  definida por  $\mathbf{g} \leq \mathbf{f}$  quando  $\langle \mathbf{g}, \mathbf{u} \rangle \leq \langle \mathbf{f}, \mathbf{u} \rangle$ , para todo  $\mathbf{u} \geq \mathbf{0}$ . Mostre que essa relação induzida no dual é uma relação de ordem parcial compatível com a estrutura vetorial do dual.
- 7.12.** Usando o morfismo canônico  $J : X \rightarrow X^{**}$  de um espaço vetorial  $X$  no seu bidual  $X^{**}$ , feito no **Teorema 4.1**, mostre que se  $X$  é de dimensão finita e  $S \subset X$  é um subespaço de  $X$ , então o anulador do anulador de  $S$  é o próprio  $S$ , i.e.  $S^{00} = S$ , ou, mais precisamente,  $J(S) = S^{00}$ .

- 7.13.** Considere o subespaço  $S = \text{span}\{(1, 0, 1, 2); (1, 2, 1, 0)\}$  de  $\mathbb{R}^4$  e caracterize os parâmetros  $(a_1, a_2, a_3, a_4)$  tais que os funcionais lineares  $\langle \mathbf{f}, \mathbf{u} \rangle = a_1 u_1 + a_2 u_2 + a_3 u_3 + a_4 u_4$ ,  $\mathbf{u} = (u_1, \dots, u_4)$ , estão no anulador  $S^0$ .
- 7.14.** Considere o subespaço  $S = \{p \in \mathcal{P}_n(\mathbb{R}); p(0) = 0\} = x\mathcal{P}_{n-1}(\mathbb{R})$  do espaço vetorial  $\mathcal{P}_n(\mathbb{R})$ , onde  $n \in \mathbb{N}$ . Encontre os funcionais lineares que estão no anulador  $S^0$  de  $S$ .
- 7.15.** Considere o subespaço  $S = \{A \in \mathbb{R}^{n \times n}; \text{tr}(A) = 0\}$  das matrizes quadradas de traço nulo em  $S = \mathbb{R}^{n \times n}$ . Mostre que  $S$  é um hiperplano e determine o anulador  $S^0$  de  $S$ , ou seja, todos os funcionais lineares que se anulam em  $S$ .
- 7.16.** Considere o subespaço  $S = \{A \in \mathbb{R}^{2 \times 2}; A^{\text{tr}} = A\}$  das matrizes simétricas em  $X = \mathbb{R}^{2 \times 2}$ . Mostre que  $S$  é um hiperplano e determine o anulador  $S^0$  de  $S$ , ou seja, todos os funcionais lineares que se anulam em  $S$ .
- 7.17.** Considere o subespaço  $S = \{A \in \mathbb{R}^{3 \times 3}; A^{\text{tr}} = A\}$  das matrizes simétricas em  $X = \mathbb{R}^{3 \times 3}$ . Qual a dimensão de  $S$ ? Qual a dimensão do anulador  $S^0$  de  $S$ ? Ache uma base para  $S^0$ .

## CAPÍTULO 3

### Transformações Lineares

Neste capítulo, vamos relembrar o conceito de transformações lineares, ver alguns exemplos e ver algumas de suas propriedades mais fundamentais.

#### 1. Fundamentos

**1.1. Definição de transformação linear.** A essência da Álgebra Linear é o estudo de transformações lineares, ou seja, que preservam a estrutura de linearidade inerente aos espaços vetoriais. De fato, os espaços vetoriais foram abstraídos dessa forma justamente por serem as estruturas fundamentais para o estudo dessas transformações. Em cima desses tipos de espaços, é natural falarmos de transformações lineares, conforme definidas a seguir.

DEFINIÇÃO 1.1. *Sejam  $X$  e  $Y$  dois espaços vetoriais sobre um mesmo corpo  $\mathbb{K}$ . Uma **transformação linear** de  $X$  em  $Y$  é uma função  $T : X \rightarrow Y$  de  $X$  em  $Y$  que é aditiva e homogênea, i.e. satisfaz*

- (i)  $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ , para todo  $\mathbf{u}, \mathbf{v} \in X$ ; e
- (ii)  $T(\lambda \mathbf{u}) = \lambda T(\mathbf{u})$ , para quaisquer  $\mathbf{u} \in X$  e  $\lambda \in \mathbb{K}$ .

Quando o domínio e o contra-domínio coincidem, denotamos a transformação linear por *operador linear*. Questões específicas de operadores lineares são discutidas a partir do [Capítulo 4](#)

**1.2. Alguns conceitos essenciais.** Assim como em teoria de conjuntos, definimos alguns conjuntos especiais associados a uma transformação linear.

DEFINIÇÃO 1.2 (Domínio e contra-domínio). *Dada uma transformação linear  $T : X \rightarrow Y$  entre dois espaços vetoriais  $X$  e  $Y$  sobre um mesmo corpo, dizemos que  $X$  é o **domínio** de  $T$  e que  $Y$  é o **contra-domínio** de  $T$ .*

DEFINIÇÃO 1.3 (Núcleo). *Dada uma transformação linear  $T : X \rightarrow Y$  entre dois espaços vetoriais  $X$  e  $Y$  sobre um mesmo corpo, definimos o **núcleo** de  $T$ , denotado por  $\ker(T)$ , como sendo o conjunto*

$$\ker(T) = \{\mathbf{u} \in X; T(\mathbf{u}) = \mathbf{0}\},$$

onde, naturalmente, o elemento  $\mathbf{0}$  à direita é o elemento de  $Y$  que é neutro para a adição vetorial.

DEFINIÇÃO 1.4 (Imagem). Dada uma transformação linear  $\mathbb{T} : X \rightarrow Y$  entre dois espaços vetoriais  $X$  e  $Y$  sobre um mesmo corpo, definimos a **imagem** de  $\mathbb{T}$ , denotada por  $\text{Im}(\mathbb{T})$ , como sendo o conjunto

$$\text{Im}(\mathbb{T}) = \{\mathbf{v} \in Y; \exists \mathbf{u} \in X, \mathbb{T}(\mathbf{u}) = \mathbf{v}\} = \{\mathbb{T}(\mathbf{u}) \in Y; \mathbf{u} \in X\}.$$

TEOREMA 1.1. Seja  $\mathbb{T} : X \rightarrow Y$  uma transformação linear entre dois espaços vetoriais  $X$  e  $Y$  sobre um mesmo corpo. O núcleo  $\ker(\mathbb{T})$  de  $\mathbb{T}$  é um subespaço vetorial do domínio  $X$ , enquanto que a imagem  $\text{Im}(\mathbb{T})$  é um subespaço vetorial do contra-domínio  $Y$ .

DEMONSTRAÇÃO. As demonstrações seguem imediatamente da propriedade de linearidade de  $\mathbb{T}$ , ou seja, pelo fato de  $\mathbb{T}$  preservar as operações de adição e multiplicação por escalar.  $\square$

DEFINIÇÃO 1.5 (Posto). Seja  $\mathbb{T} : X \rightarrow Y$  uma transformação linear entre dois espaços vetoriais  $X$  e  $Y$  sobre um mesmo corpo. O **posto** de  $\mathbb{T}$  é a dimensão da sua imagem.

### 1.3. Exemplos.

EXEMPLO 1.1 (Transformação linear associada a uma matriz). Com  $X = \mathbb{R}^n$  e  $Y = \mathbb{R}^m$ , sendo  $n, m \in \mathbb{N}$ , cada matriz  $A \in \mathbb{R}^{m \times n}$  gera, naturalmente, uma transformação linear associada  $\mathbb{T}_A : \mathbb{R}^n \rightarrow \mathbb{R}^m$  definida pela multiplicação da matriz  $A = (a_{ij})_{i,j=1}^{i=m,j=n}$  por um vetor  $\mathbf{u} = (u_j)_{j=1,\dots,n} \in \mathbb{R}^n$  definida por

$$\mathbb{T}_A(\mathbf{u}) = A\mathbf{u} = \left( \sum_{j=1}^n a_{ij}u_j \right)_{i=1,\dots,m} = \begin{pmatrix} a_{11}u_1 + \cdots + a_{1n}u_n \\ a_{21}u_1 + \cdots + a_{2n}u_n \\ \vdots \\ a_{m1}u_1 + \cdots + a_{mn}u_n \end{pmatrix}.$$

EXEMPLO 1.2 (Rotação). A transformação linear  $\mathbb{T} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  definida por  $\mathbb{T}(x, y) = (-y, x)$  é a matriz de rotação de um vetor  $\mathbf{u} = (x, y)$  por  $90^\circ$ , no sentido trigonométrico, ou anti-horário. Ela é representada pela matriz

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

Mais geralmente,  $\mathbb{T}_\theta(x, y) = (\cos(\theta)x - \text{sen}(\theta)y, \text{sen}(\theta)x + \cos(\theta)y)$ , representada pela matriz

$$R_\theta = \begin{bmatrix} \cos(\theta) & -\text{sen}(\theta) \\ \text{sen}(\theta) & \cos(\theta) \end{bmatrix},$$

é uma rotação de um ângulo  $\theta$  qualquer, também no sentido trigonométrico.

EXEMPLO 1.3 (Representação polinomial). A transformação  $J : \mathcal{P}_n(\mathbb{R}) \rightarrow \mathbb{R}^{n+1}$  definida por

$$J(p) = \left( p(0), \frac{dp}{dx}(0), \frac{1}{2} \frac{d^2p}{dx^2}(0), \dots, \frac{1}{n!} \frac{d^n p}{dx^n}(0) \right) \in \mathbb{R}^n, \quad \forall p \in \mathcal{P}_n(\mathbb{R}),$$

é a transformação linear que leva um polinômio  $p = p(x)$  de grau no máximo  $n$  nos coeficientes  $J(p) = (a_0, \dots, a_n)$  de sua representação canônica

$$p(x) = a_0 + a_1x + \dots + a_nx^n,$$

visto que  $a_0 = p(0)$  e que cada  $a_j$ ,  $j = 1, \dots, n$  é exatamente a derivada de ordem  $j$  de  $p$  dividida por  $j$  fatorial. Essa transformação linear é um isomorfismo entre  $\mathcal{P}_n(\mathbb{R})$  e  $\mathbb{R}^{n+1}$ . Idem para  $J : \mathcal{P}_n(\mathbb{C}) \rightarrow \mathbb{C}^{n+1}$ , onde as derivadas, agora, são derivadas complexas,

$$J(p) = \left( p(0), \frac{dp}{dz}(0), \frac{1}{2} \frac{d^2p}{dz^2}(0), \dots, \frac{1}{n!} \frac{d^n p}{dz^n}(0) \right) \in \mathbb{C}^n, \quad \forall p \in \mathcal{P}_n(\mathbb{C}).$$

EXEMPLO 1.4 (Isomorfismos). Vimos, no [Exemplo 1.3](#), um exemplo de transformação linear que é um isomorfismo. De fato, todo isomorfismo  $J : X \rightarrow Y$  entre dois espaços vetoriais  $X$  e  $Y$  sobre um mesmo corpo é, por definição, um exemplo de transformação linear. Mais ainda, a inversa  $J^{-1} : Y \rightarrow X$  é, também, uma transformação linear, agora de  $Y$  em  $X$ . Verifique ([Exercício 6.1](#))! Veja mais sobre isso na [Seção 3.2](#).

EXEMPLO 1.5 (Diferencial). Assim como no caso de funcionais, um exemplo fundamental é o de diferencial de uma função (usualmente não linear) entre dois espaços vetoriais sobre o mesmo corpo, quando munidos de normas. Mais precisamente, se  $F : X \rightarrow Y$  é uma função entre dois espaços vetoriais normados sobre o mesmo corpo, com as normas denotadas por  $\|\cdot\|_X$  e  $\|\cdot\|_Y$ , respectivamente, dizemos que  $F$  é **diferenciável** em um ponto  $\mathbf{u} \in X$ , quando existe uma transformação linear  $L : X \rightarrow Y$ , que é a chamada **diferencial** de  $f$  no ponto  $\mathbf{u}$  e denotada por  $DF(\mathbf{u}) = L$ , que satisfaz

$$\frac{\|F(\mathbf{u} + \mathbf{h}) - F(\mathbf{u}) - (DF(\mathbf{u}))(\mathbf{h})\|_Y}{\|\mathbf{h}\|_X} \rightarrow 0, \quad \text{quando } \|\mathbf{h}\|_X \rightarrow \mathbf{0}, \mathbf{h} \neq \mathbf{0}.$$

Quando  $\mathbf{u} = (x_1, \dots, x_n) \in X = \mathbb{R}^n$  e  $F(\mathbf{u}) = (f_1(x_1, \dots, x_n), \dots, f_m(x_1, \dots, x_n)) \in Y = \mathbb{R}^m$ ,  $n, m \in \mathbb{N}$ , então a diferencial calculada em um vetor  $\mathbf{h}$  pode ser *representada* por uma multiplicação matricial  $DF(\mathbf{u})\mathbf{h}$ , onde  $DF(\mathbf{u}) = (\partial_{x_j} f_i(x_1, \dots, x_n))_{i,j=1}^{i=m, j=n}$ , que é, então, nesse caso, tomada como a definição de diferencial de  $F$ .

EXEMPLO 1.6. Dadas duas variedades diferenciáveis  $\mathcal{M}$ ,  $\mathcal{N}$  (veja [Exemplo 1.14](#)) e uma função diferenciável  $\mathbf{F} : \mathcal{M} \rightarrow \mathcal{N}$ , a diferencial  $D\mathbf{F}(\mathbf{p}) : T_{\mathbf{p}}\mathcal{M} \rightarrow T_{\mathbf{F}(\mathbf{p})}\mathcal{N}$  de  $\mathbf{F}$  no ponto  $\mathbf{p}$  é uma transformação linear entre a variedade tangente a  $\mathcal{M}$  no ponto  $\mathbf{p}$ , que é o espaço vetorial  $T_{\mathbf{p}}\mathcal{M}$ , e a variedade tangente a  $\mathcal{N}$  no ponto da imagem  $\mathbf{F}(\mathbf{p})$ , que é o espaço vetorial  $T_{\mathbf{F}(\mathbf{p})}\mathcal{N}$ .

EXEMPLO 1.7 (Redes neurais). Redes neurais podem ter diversas formas e não são transformações lineares, mas boa parte da sua estrutura depende de transformações lineares. As redes pró-alimentadas densas, por exemplo, são composições de transformações lineares com funções escalares chamadas de *funções de ativação*, simulando o comportamento de neurônios, tendo, em geral, essencialmente dois estados, “ativado” ou “não”, ou estados variando de intensidade. Podemos escrever, mais precisamente, uma rede pró-alimentada na forma

$$h(\mathbf{x}) = \sigma_k \circ \mathbf{W}_k \circ \sigma_{k-1} \circ \mathbf{W}_{k-1} \circ \cdots \circ \sigma_1 \circ \mathbf{W}_1(\mathbf{x}),$$

onde cada  $\mathbf{W}_j : \mathbb{R}^{n_{j-1}} \rightarrow \mathbb{R}^{n_j}$  é uma transformação linear, entre espaços de dimensão  $n_{j-1}$  e  $n_j$ ; cada  $j = 0, \dots, k$  representa uma camada da rede neural; cada  $n_j$  é a dimensão, ou largura, da camada, indicando o “número de neurônios” em cada camada, que nada mais são do que variáveis reais; e cada  $\sigma_j : \mathbb{R} \rightarrow \mathbb{R}$  é uma função de ativação, estendida para agir em cada espaço  $\mathbb{R}^{n_j}$  via  $\sigma(\mathbf{u}) = (\sigma(x_j))_{j=1, \dots, n_j}$ , para  $\mathbf{u} = (x_1, \dots, x_{n_j}) \in \mathbb{R}^{n_j}$ . A primeira camada, correspondendo a  $j = 0$ , é a entrada da rede, enquanto que a última camada, correspondendo a  $j = n$ , é a saída. As camadas intermediárias,  $j = 1, \dots, n - 1$ , são as chamadas camadas ocultas. Veja [Figura 1.1](#).

EXEMPLO 1.8 (Convolução - molificação). A convolução entre duas funções  $f, g : \mathbb{R}^n \rightarrow \mathbb{R}$  é a operação que leva essas duas funções na função  $f \star g$  definida por

$$(f \star g)(\mathbf{x}) = \int_{\mathbb{R}^n} f(\mathbf{y})g(\mathbf{x} - \mathbf{y}) \, d\mathbf{y} = \int_{\mathbb{R}^n} f(\mathbf{x} - \mathbf{y})g(\mathbf{y}) \, d\mathbf{y},$$

quando a integral está bem definida. Isso acontece, por exemplo, para  $f \in L^p(\mathbb{R})$  e  $g \in L^q(\mathbb{R})$ , onde  $1 \leq p, q \leq \infty$  satisfazem

$$\frac{1}{p} + \frac{1}{q} = 1.$$

A convolução é bilinear em  $f$  e  $g$ . Uma convolução muito usada é a entre uma função  $f$  e um *molificador* suave  $\rho_\varepsilon$  da forma

$$\rho_\varepsilon(\mathbf{x}) = \frac{1}{\varepsilon^n} \rho\left(\frac{\mathbf{x}}{\varepsilon}\right),$$

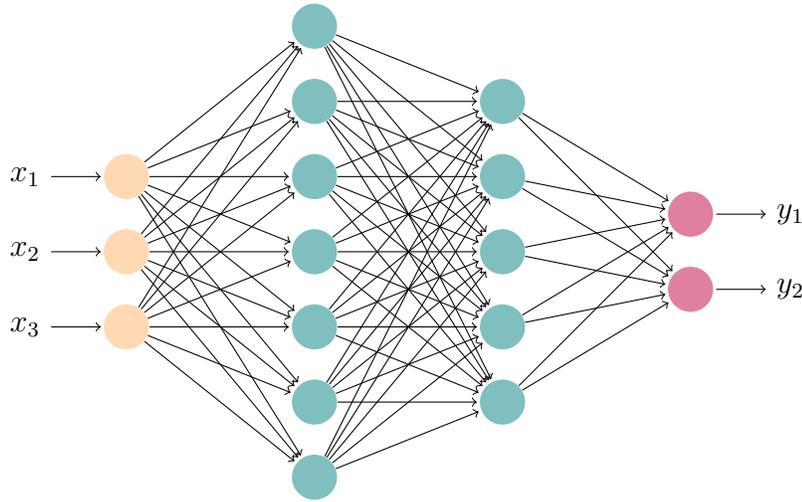


FIGURA 1.1. Ilustração de uma rede neural densa, com três entradas, duas saídas e duas camadas ocultas, onde  $k = 3$ ,  $n_0 = 3$ ,  $n_1 = 7$ ,  $n_2 = 5$ ,  $n_3 = 2$ .

onde  $\rho \in \mathcal{C}_c^\infty(\mathbb{R}^n)$  (i.e. funções de suporte compacto, diferenciáveis de qualquer ordem e com diferenciais contínuas) e

$$\int_{\mathbb{R}^n} \rho(\mathbf{x}) \, d\mathbf{x} = 1.$$

Nesse caso, sendo  $f \in L^p(\mathbb{R}^n)$ ,  $1 \leq p < \infty$ , a molificação  $\rho_\varepsilon \star f$  é uma função ainda em  $L^p(\mathbb{R}^n)$ , de classe  $\mathcal{C}^\infty$ , e que converge para  $f$  em  $L^p(\mathbb{R}^n)$ . O mapeamento entre  $f$  e  $\mathbb{T}_\varepsilon(f) = f_\varepsilon = \rho_\varepsilon \star f$  é uma transformação linear de  $L^p(\mathbb{R}^n)$  em  $L^p(\mathbb{R}^n) \cap \mathcal{C}^\infty(\mathbb{R}^n)$ .

EXEMPLO 1.9 (Equações diferenciais lineares). Uma equação diferencial linear

$$a(t) \frac{d^2 x(t)}{dt^2} + b(t) \frac{dx(t)}{dt} + c(t)x(t) = f(t)$$

pode ser interpretada como uma equação linear

$$\mathbf{A}x = f,$$

para uma incógnita  $x \in X = \mathcal{C}^2(\mathbb{R})$ , com  $f \in Y = \mathcal{C}(\mathbb{R})$  dado e  $\mathbf{A} : X \rightarrow Y$  definido por

$$\mathbf{A} = a \frac{d^2}{dt^2} + b \frac{d}{dt} + c,$$

ou seja, para cada  $x = x(t) \in X$ , temos  $Ax \in Y$  definido por

$$(Ax)(t) = a(t)\frac{d^2x(t)}{dt^2} + b(t)\frac{dx(t)}{dt} + c(t),$$

onde  $a, b, c \in X$  também são conhecidos. O problema homogêneo, i.e.

$$a(t)\frac{d^2x(t)}{dt^2} + b(t)\frac{dx(t)}{dt} + c(t)x(t) = 0$$

é o núcleo  $\ker(A)$  da transformação  $A$ . A solução geral da equação não-homogênea é um subespaço afim

$$x = x_h + x_p,$$

onde  $x_h \in \ker(A)$  e  $x_p = x_p(t)$  é uma solução particular, ou seja, um representante qualquer do subespaço  $X/\ker(A)$ .

EXEMPLO 1.10 (Sistema de resposta ao impulso). Um sistema de resposta ao impulso (*Impulse Response - IR*) é uma transformação que recebe um breve *impulso* (como uma delta de Dirac em um determinado instante, no caso contínuo, ou um Delta de Kronecker, no caso discreto) e retorna um determinado sinal de resposta. Em teoria, a resposta pode ser linear ou não-linear, mas o caso linear é de extrema importância prática e muito usada no processamento de sinais de áudio, de imagem e de vídeo. No caso de sinais de áudio, por exemplo, os sistemas de resposta de impulso são usados na simulação de amplificadores e ambientes para se produzir efeitos de guitarra e de outros aparelhos musicais. Primeiro, grava-se, em estúdio, a resposta, a um determinado sinal de áudio, de um amplificador real, em um ambiente de controlado de estúdio ou em algum ambiente de exibição musical. Em seguida, o espectro dessa resposta é analisado, via transformada de Laplace, como uma *assinatura* do que está se querendo capturar. A partir dessa análise, uma transformação linear, na forma de convolução, é gerada. Posteriormente, essa convolução pode ser aplicada, então, a sinais de guitarra ou de outros instrumentos musicais, para simular o efeito real de se usar aquele determinado amplificador ou ambiente, e assim por diante.

EXEMPLO 1.11 (Tratamento de imagem). Em tratamento de imagem, vários filtros podem ser utilizados, para gerar diversos efeitos. Nem todos são lineares, mas uma boa gama deles tem, sim, a forma linear. Por exemplo, o filtro Gaussiano aplica uma (forma discretizada) de convolução com uma função Gaussiana, visando suavizar uma imagem, ou uma determinada região de uma imagem.

EXEMPLO 1.12 (Transformada de Fourier). A transformada de Fourier pode ser vista, inicialmente, como uma transformação linear  $F : L^1(\mathbb{R}^n, \mathbb{C}) \rightarrow L^\infty(\mathbb{R}^n, \mathbb{C})$ , que

leva uma função integrável a Lebesgue  $f : \mathbb{R}^n \rightarrow \mathbb{C}$  em uma função essencialmente limitada  $\hat{f} = F(f) : \mathbb{R}^n \rightarrow \mathbb{C}$  dada por

$$\hat{f}(\boldsymbol{\omega}) = F(f)(\boldsymbol{\omega}) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} f(\mathbf{x}) e^{-i\boldsymbol{\omega} \cdot \mathbf{x}} \, d\mathbf{x}, \quad \boldsymbol{\omega} \in \mathbb{R}^n, \quad (1.1)$$

onde  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n) \in \mathbb{R}^n$  representa a *frequência angular*, nas diversas direções. Há outras formulações, por exemplo, definindo a transformada de Fourier como

$$\int_{\mathbb{R}^n} f(\mathbf{x}) e^{-i2\pi\xi \cdot \mathbf{z}} \, d\mathbf{x},$$

agora com a *frequência*  $\xi = 2\pi\boldsymbol{\omega}$ , mais diretamente relacionada à série de Fourier clássica. Com a formulação (1.1), no entanto, a transformada inversa  $F^{-1}$  tem uma forma mais parecida com a transformada direta, por ter o mesmo fator multiplicativo. Ainda com essa formulação, a relação entre a derivada, no espaço físico, e a multiplicação algébrica, no espaço de frequências, toma a forma

$$F(\partial_{x_i}^k f) = (i\xi)^k F(f).$$

**1.4. Transformação linear definida a partir da ação em uma base.** Graças a linearidade, podemos definir uma transformação linear apenas definindo o que acontece com os vetores de uma base.

**TEOREMA 1.2.** *Sejam  $X$  e  $Y$  espaços vetoriais de dimensão finita sobre um mesmo corpo  $\mathbb{K}$  e seja  $\mathfrak{a} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  uma base de  $X$ . Sejam  $\mathbf{v}_1, \dots, \mathbf{v}_n \in Y$  vetores quaisquer em  $Y$ . Então existe uma única transformação linear  $T : X \rightarrow Y$  tal que  $T(\mathbf{w}_i) = \mathbf{v}_i$ , para cada  $i = 1, \dots, n$ . Nesse caso, dizemos que  $T$  é obtida por **extensão linear**, a partir da ação em  $\mathfrak{a}$ .*

**DEMONSTRAÇÃO.** Dado um  $\mathbf{u} \in X$  qualquer, podemos escrever

$$\mathbf{u} = \langle \mathfrak{f}_1, \mathbf{u} \rangle \mathbf{w}_1 + \dots + \langle \mathfrak{f}_n, \mathbf{u} \rangle \mathbf{w}_n,$$

onde  $\mathfrak{a}^* = \{\mathfrak{f}_1, \dots, \mathfrak{f}_n\}$  é a base dual a  $\mathfrak{a}$ . A partir dessa fórmula, podemos definir  $T : X \rightarrow Y$  por

$$T(\mathbf{u}) = \langle \mathfrak{f}_1, \mathbf{u} \rangle \mathbf{v}_1 + \dots + \langle \mathfrak{f}_n, \mathbf{u} \rangle \mathbf{v}_n \in Y.$$

Se  $\mathbf{v} \in X$  é outro vetor e  $\lambda \in \mathbb{K}$ , obtemos,

$$\begin{aligned} T(\mathbf{u} + \lambda\mathbf{v}) &= \langle \mathfrak{f}_1, \mathbf{u} + \lambda\mathbf{v} \rangle \mathbf{v}_1 + \dots + \langle \mathfrak{f}_n, \mathbf{u} + \lambda\mathbf{v} \rangle \mathbf{v}_n \\ &= \langle \mathfrak{f}_1, \mathbf{u} \rangle \mathbf{v}_1 + \dots + \langle \mathfrak{f}_n, \mathbf{u} \rangle \mathbf{v}_n + \lambda \langle \mathfrak{f}_1, \mathbf{v} \rangle \mathbf{v}_1 + \dots + \lambda \langle \mathfrak{f}_n, \mathbf{v} \rangle \mathbf{v}_n \\ &= T(\mathbf{u}) + \lambda T(\mathbf{v}), \end{aligned}$$

mostrando que  $T$  é linear. Como não há outra forma de escrever um vetor  $\mathbf{u}$  como combinação linear dos vetores da base e a ação nesses vetores da base está bem definida, essa é a única transformação linear possível.  $\square$

EXEMPLO 1.13. Em  $\mathbb{R}^2$ , considerando os mapeamentos

$$\mathbf{e}_1 \rightarrow \mathbf{e}_2, \quad \mathbf{e}_2 \rightarrow -\mathbf{e}_1,$$

vemos que a transformação obtida por extensão linear desses mapeamentos é a rotação de 90 graus no sentido trigonométrico, com

$$x\mathbf{e}_1 + y\mathbf{e}_2 \rightarrow x(\mathbf{e}_2) + y(-\mathbf{e}_1) = -y\mathbf{e}_1 + x\mathbf{e}_2,$$

ou seja,  $T(x, y) = (-y, x)$ .

EXEMPLO 1.14. Em  $\mathbb{R}^3$ , a transformação linear obtida por extensão linear de

$$\mathbf{e}_1 \rightarrow \mathbf{e}_1, \quad \mathbf{e}_2 \rightarrow \mathbf{e}_2, \quad \mathbf{e}_3 \rightarrow -\mathbf{e}_3$$

é a reflexão de  $xyz$  em relação ao plano  $xy$ , dada mais explicitamente por  $T(x, y, z) = (x, y, -z)$ .

EXEMPLO 1.15. De  $\mathbb{R}^3$  para  $\mathbb{R}^2$ , considerando

$$(1, 0, 0) \rightarrow (1, 1), \quad (0, 1, 0) \rightarrow (1, 1), \quad (0, 0, 1) \rightarrow (0, 0)$$

obtemos, por extensão linear, a transformação  $T(x, y, z) = (x + y, x + y)$ .

## 2. O Teorema do Núcleo e da Imagem

Esse é um teorema fundamental em Álgebra Linear. É um resultado de “contabilidade”: para onde vão as “dimensões” do espaço domínio? Parte pertence ao núcleo da transformação e é anulado por ela, sendo levado no elemento neutro do contra-domínio. A parte restante vai para além e forma a imagem da transformação. Ou seja, a dimensão do espaço domínio é repartida entre a dimensão do núcleo e a dimensão da imagem. A imagem é, de fato, isomorfa ao espaço quociente do núcleo.

### 2.1. Enunciado e demonstração.

TEOREMA 2.1 (do Núcleo e da Imagem). *Seja  $T : X \rightarrow Y$  uma transformação linear entre dois espaços vetoriais  $X$  e  $Y$  sobre um mesmo corpo, com  $X$  de dimensão finita. Então*

$$\dim(X) = \dim(\ker(T)) + \dim(\text{Im}(T)). \quad (2.1)$$

DEMONSTRAÇÃO. A demonstração clássica parte de uma base  $\{\mathbf{w}_1, \dots, \mathbf{w}_k\}$  para  $\ker(\mathbb{T})$  e a completa até uma base  $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  de  $X$ , no caso em que  $1 \leq k < n$ , onde  $k = \dim(\ker(\mathbb{T}))$  e  $n = \dim(X)$ . Todo  $\mathbf{u} \in X$  se escreve de forma única como

$$\mathbf{u} = x_1 \mathbf{w}_1 + \dots + x_n \mathbf{w}_n.$$

Aplicando  $\mathbb{T}$  e considerando que  $\mathbb{T}(\mathbf{w}_i) = \mathbf{0}$  para todo  $i = 1, \dots, k$ , obtemos

$$\mathbb{T}(\mathbf{u}) = x_{k+1} \mathbb{T}(\mathbf{w}_{k+1}) + \dots + x_n \mathbb{T}(\mathbf{w}_n).$$

Portanto, a imagem  $\text{Im}(\mathbb{T})$  é gerado pelos vetores  $\{\mathbb{T}(\mathbf{w}_{k+1}), \dots, \mathbb{T}(\mathbf{w}_n)\}$ . Para ver que eles são linearmente independentes, suponha que

$$\lambda_{k+1} \mathbb{T}(\mathbf{w}_{k+1}) + \dots + \lambda_n \mathbb{T}(\mathbf{w}_n) = \mathbf{0}.$$

Por linearidade, temos

$$\mathbb{T}(\lambda_{k+1} \mathbf{w}_{k+1} + \dots + \lambda_n \mathbf{w}_n) = \mathbf{0}.$$

Ou seja,  $\lambda_{k+1} \mathbf{w}_{k+1} + \dots + \lambda_n \mathbf{w}_n \in \ker(\mathbb{T})$ . Como  $\{\mathbf{w}_1, \dots, \mathbf{w}_k\}$  é base de  $\ker(\mathbb{T})$ , então, aquele vetor é combinação linear destes, ou seja, existem  $\lambda_1, \dots, \lambda_k \in \mathbb{K}$  tais que

$$\lambda_{k+1} \mathbf{w}_{k+1} + \dots + \lambda_n \mathbf{w}_n = \lambda_1 \mathbf{w}_1 + \dots + \lambda_k \mathbf{w}_k,$$

ou seja,

$$\lambda_1 \mathbf{w}_1 + \dots + \lambda_k \mathbf{w}_k - \lambda_{k+1} \mathbf{w}_{k+1} - \dots - \lambda_n \mathbf{w}_n = \mathbf{0}.$$

Mas como  $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  é uma base de  $X$ , então esses vetores são linearmente independentes, de modo que, necessariamente, os coeficientes devem todos se anular, i.e.  $\lambda_1 = \dots = \lambda_n = 0$ . Em particular, isso prova que  $\{\mathbb{T}(\mathbf{w}_{k+1}), \dots, \mathbb{T}(\mathbf{w}_n)\}$  forma um conjunto LI. Como já vimos que eles geram  $\text{Im}(\mathbb{T})$ , então eles formam uma base para essa imagem. Ou seja,  $\dim(\text{Im}(\mathbb{T})) = n - k$ . Assim,

$$\dim(X) = n = k + (n - k) = \dim(\ker(\mathbb{T})) + \dim(\text{Im}(\mathbb{T})).$$

Faltam os casos em que um dos subespaços é trivial. Além do caso em que o próprio  $X$  é trivial e, assim, ambos os subespaços são triviais. Começando com o caso em que  $X = \{\mathbf{0}\}$ , então necessariamente  $\ker(\mathbb{T}) = \{\mathbf{0}\}$  e  $\text{Im}(\mathbb{T}) = \{\mathbf{0}\}$ , de modo que todas as dimensões são nulas e o resultado é trivialmente válido.

Caso  $\ker(\mathbb{T}) = \{\mathbf{0}\}$  e  $\dim(X) = n \in \mathbb{N}$ , então não partimos de uma base de  $\ker(\mathbb{T})$ , já que tal não existe. Tomamos, então, uma base qualquer  $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  de  $X$ . Obtemos, de forma semelhante à feita acima, que  $\{\mathbb{T}(\mathbf{w}_1), \dots, \mathbb{T}(\mathbf{w}_n)\}$  é base de  $\text{Im}(\mathbb{T})$ , de modo que  $\dim(X) = \dim(\text{Im}(\mathbb{T}))$  e  $\dim(\ker(\mathbb{T})) = 0$ , satisfazendo também a igualdade (2.1).

Finalmente, caso  $\dim(\ker(\mathbb{T})) = n = \dim(X)$ , então partimos de uma base de  $\ker(\mathbb{T})$  com  $n$  elementos, digamos  $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ . Nesse caso, essa também é uma base de  $X$  e todo  $\mathbf{u} \in X$  se escreve como combinação linear de elementos do núcleo, sendo,

FIGURA 2.1. Ilustração do Teorema do Núcleo e da Imagem através do isomorfismo entre  $X/\ker(\mathbb{T})$  e  $\text{Im}(\mathbb{T})$ , para uma transformação linear  $\mathbb{T} : X \rightarrow Y$  entre dois espaços vetoriais sobre o mesmo corpo.

portanto, também parte do núcleo, ou seja, a imagem é necessariamente  $\text{Im}(\mathbb{T}) = \{\mathbf{0}\}$ . Assim,  $\dim(\text{Im}(\mathbb{T})) = 0$  e (2.1) também é válida.

Isso completa a demonstração.  $\square$

OBSERVAÇÃO 2.1. Uma outra demonstração do Teorema 2.1, bastante elegante, que aparece, por exemplo, no livro do Lax [8], usa espaço quociente. Além de elegante, é também ilustrativa. Veja Figura 2.1. Passa por mostrar que  $X/\ker(\mathbb{T})$  é isomorfo a  $\text{Im}(\mathbb{T})$  via isomorfismo  $J([\mathbf{u}]) = J(\mathbf{u} + \ker(\mathbb{T})) = \mathbb{T}(\mathbf{u})$ , independente do representante da classe  $[\mathbf{u}] \in X/\ker(\mathbb{T})$ . Assim, usando (2.1), obtemos

$$\dim(\text{Im}(\mathbb{T})) = \dim(X/\ker(\mathbb{T})) = \dim(X) - \dim(\ker(\mathbb{T})).$$

Deixamos a verificação de que  $X/\ker(\mathbb{T})$  é isomorfo a  $\text{Im}(\mathbb{T})$  para o leitor (Exercício 6.3)

## 2.2. Consequências imediatas.

TEOREMA 2.2. *Se  $\dim(Y) < \dim(X)$ , então necessariamente  $\dim(\ker(\mathbb{T})) \geq \dim(X) - \dim(Y) \geq 1$ .*

DEMONSTRAÇÃO. Como  $\text{Im}(\mathbb{T}) \subset Y$ , o resultado segue imediatamente de

$$\dim(\ker(\mathbb{T})) = \dim(X) - \dim(\text{Im}(\mathbb{T})) \geq \dim(X) - \dim(Y) \geq 1.$$

$\square$

TEOREMA 2.3. *Se  $\dim(Y) = \dim(X)$ , então  $\ker(\mathbb{T}) = \{\mathbf{0}\}$  se, e somente se,  $\text{Im}(\mathbb{T}) = Y$  se, e somente se,  $\mathbb{T}$  é uma bijeção, ou seja,  $\mathbb{T}$  é injetivo se, e somente se, é sobrejetivo e se, e somente se, é bijetivo.*

DEMONSTRAÇÃO. Segue de Equation (2.1) e da hipótese do teorema que

$$\dim(\ker(\mathbb{T})) = \dim(X) - \dim(\text{Im}(\mathbb{T})) = \dim(Y) - \dim(\text{Im}(\mathbb{T})).$$

Logo,

$$\ker(\mathbb{T}) = \{\mathbf{0}\} \Leftrightarrow \dim(\ker(\mathbb{T})) = 0 \Leftrightarrow \dim(\text{Im}(\mathbb{T})) = \dim(Y) \Leftrightarrow \text{Im}(\mathbb{T}) = Y,$$

ou seja,  $\mathbb{T}$  é injetivo se, e somente se, é sobrejetivo e, portanto, se, e somente se, é bijetivo.  $\square$

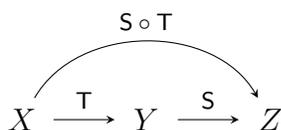


FIGURA 3.1. Ilustração da composição  $T \circ S : X \rightarrow Z$  de transformações  $T : X \rightarrow Y$  e  $S : Y \rightarrow Z$ .

### 3. O espaço das transformações lineares

**3.1. O espaço vetorial das transformações lineares.** Dados dois espaços vetoriais  $X$  e  $Y$  sobre um mesmo corpo, podemos considerar o conjunto de *todas* as transformações lineares de  $X$  em  $Y$ . Nesse espaço, podemos estender, de forma natural, as operações de adição e multiplicação por escalar dos espaços vetoriais para esse conjunto de transformações, fazendo deste um espaço vetorial, também.

DEFINIÇÃO 3.1 (Espaço vetorial das transformações lineares). *Dados dois espaços vetoriais  $X$  e  $Y$  sobre um mesmo corpo  $\mathbb{K}$ , denotamos o espaço de todas as transformações lineares de  $X$  em  $Y$  por*

$$\mathcal{L}(X, Y) = \{T : X \rightarrow Y; T \text{ linear}\}.$$

*Esse espaço  $\mathcal{L}(X, Y)$  é um espaço vetorial sobre  $\mathbb{K}$  quando munido das operações de adição e multiplicação por escalar definidas pontualmente, i.e. para todo  $T, S \in \mathcal{L}(X, Y)$ , a adição  $T + S \in \mathcal{L}(X, Y)$  é a transformação linear de  $X$  em  $Y$  definida por*

$$(T + S)(\mathbf{u}) = T(\mathbf{u}) + S(\mathbf{u}), \quad \forall \mathbf{u} \in X,$$

*e, para todo  $\lambda \in \mathbb{K}$  e todo  $T \in \mathcal{L}(X, Y)$ , a multiplicação por escalar  $\lambda T \in \mathcal{L}(X, Y)$  é a transformação linear de  $X$  em  $Y$  definida dada por*

$$(\lambda T)(\mathbf{u}) = \lambda T(\mathbf{u}), \quad \forall \mathbf{u} \in X.$$

Uma propriedade extra é a de que as transformações lineares podem ser compostas umas com as outras, desde que os espaços sejam compatíveis.

DEFINIÇÃO 3.2. *Sejam  $X, Y$  e  $Z$  espaços vetoriais sobre um mesmo corpo e sejam  $T \in \mathcal{L}(X, Y)$ ,  $S \in \mathcal{L}(Y, Z)$  transformações lineares. Então a **composição**  $S \circ T : X \rightarrow Z$  de  $T$  com  $S$ , também denotada simplesmente por  $ST$ , é definida por*

$$(ST)(\mathbf{u}) = (S \circ T)(\mathbf{u}) = S(T(\mathbf{u})) \in Z, \quad \forall \mathbf{u} \in X.$$

É fácil ver que a composição é uma transformação linear.

PROPOSIÇÃO 3.1. *Sejam  $X, Y$  e  $Z$  espaços vetoriais sobre um mesmo corpo e sejam  $T \in \mathcal{L}(X, Y)$ ,  $S \in \mathcal{L}(Y, Z)$ . Então  $S \circ T \in \mathcal{L}(X, Z)$ , ou seja, a composição é, também, uma transformação linear.*

DEMONSTRAÇÃO. Basta observar que

$$\begin{aligned} (S \circ T)(\mathbf{u} + \lambda \mathbf{v}) &= S(T(\mathbf{u} + \lambda \mathbf{v})) = S(T(\mathbf{u}) + \lambda T(\mathbf{v})) \\ &= S(T(\mathbf{u})) + \lambda S(T(\mathbf{v})) = (S \circ T)(\mathbf{u}) + \lambda (S \circ T)(\mathbf{v}), \end{aligned}$$

para todos  $\mathbf{u}, \mathbf{v} \in X$  e todo  $\lambda \in \mathbb{K}$ .  $\square$

Podemos combinar a composição com as operações de adição e multiplicação por escalar. Nesse caso, podemos observar que a composição é *distributiva* em relação à adição:

PROPOSIÇÃO 3.2. *Sejam  $X, Y$  e  $Z$  espaços vetoriais sobre um mesmo corpo e sejam  $T, L \in \mathcal{L}(X, Y)$  e  $R, S \in \mathcal{L}(Y, Z)$ . Então a composição é distributiva à esquerda, i.e.*

$$S \circ (T + L) = (S \circ T) + (S \circ L)$$

e distributiva à direita, i.e.

$$(S + R) \circ T = (S \circ T) + (R \circ T).$$

DEMONSTRAÇÃO. A demonstração da distributividade à direita é imediata da definição de composição, enquanto que a distributividade à esquerda usa a linearidade de  $S$ .  $\square$

Podemos, também, combinar três ou mais composições. Nesse sentido, podemos dizer que a composição é *associativa*.

PROPOSIÇÃO 3.3. *Sejam  $X, Y, Z$  e  $W$  espaços vetoriais sobre um mesmo corpo e sejam  $T \in \mathcal{L}(X, Y)$ ,  $R \in \mathcal{L}(Y, Z)$  e  $S \in \mathcal{L}(Z, W)$ . Então*

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

DEMONSTRAÇÃO. A demonstração também é imediata.  $\square$

OBSERVAÇÃO 3.1. As propriedades acima de distributividade e associatividade da composição, em relação à adição, se assemelham às propriedades da multiplicação em um anel. Esses motivos justificam a notação de composição  $S \circ T$  como um produto  $ST$ . Em relação a isso, vale ressaltar que também podemos enxergar a multiplicação de matrizes como a composição das transformações associadas a elas, justificando, novamente, a notação de produto para a composição.

OBSERVAÇÃO 3.2. Nem sempre faz sentido considerar a composição na ordem contrária também, i.e.  $T \circ S$  junto com  $S \circ T$ . Isso só faz sentido quando o domínio e o contra-domínio coincidem. Mesmo assim, nesses casos, apesar das duas composições fazerem sentido, elas podem não coincidir. Isso é discutido na [Seção 1.2](#).

OBSERVAÇÃO 3.3. No caso de duas transformações lineares associadas a matrizes (veja [Exemplo 1.1](#)), digamos  $T_A : \mathbb{K}^n \rightarrow \mathbb{K}^p$  e  $T_B : \mathbb{K}^p \rightarrow \mathbb{K}^m$ , com  $n, p, m \in \mathbb{N}$ ,  $A \in \mathbb{K}^{p \times n}$  e  $B \in \mathbb{K}^{m \times p}$ , onde  $T_A(\mathbf{u}) = A\mathbf{u}$ , para  $\mathbf{u} \in \mathbb{K}^n$  e  $T_B(\mathbf{v}) = B\mathbf{v}$ , para  $\mathbf{v} \in \mathbb{K}^p$ , a transformação composta é a transformação associada ao produto das matrizes, i.e.  $T_B \circ T_A = T_{BA}$ , onde  $T_{BA}\mathbf{u} = BA\mathbf{u}$ , para  $\mathbf{u} \in \mathbb{K}^n$ . De fato, se  $A = (a_{ij})_{i=1, \dots, m, j=1, \dots, p}$  e  $B = (b_{ij})_{i=1, \dots, p, j=1, \dots, n}$ , então, para  $\mathbf{u} = (x_1, \dots, x_n) \in \mathbb{R}^n$ ,

$$\begin{aligned} (T_B \circ T_A)(\mathbf{u}) &= T_B(T_A\mathbf{u}) \\ &= T_B \left( \left( \sum_{j=1, \dots, n} a_{kj} x_j \right)_{k=1, \dots, p} \right) \\ &= \left( \sum_{k=1, \dots, p} b_{ik} \left( \sum_{j=1, \dots, n} a_{kj} x_j \right) \right)_{i=1, \dots, m} \\ &= \left( \sum_{k=1, \dots, p} \sum_{j=1, \dots, n} b_{ik} a_{kj} \right)_{i=1, \dots, m} \\ &= T_{BA}\mathbf{u}. \end{aligned}$$

**3.2. Invertibilidade de transformações lineares.** Se  $T : X \rightarrow Y$  é uma bijeção, então existe a sua inversa  $T^{-1} : Y \rightarrow X$  que é uma transformação linear de  $Y$  em  $X$ . Essa inversa satisfaz  $T^{-1}(T(\mathbf{u})) = \mathbf{u}$ , para todo  $\mathbf{u} \in X$  e  $T(T^{-1}(\mathbf{v})) = \mathbf{v}$ , para todo  $\mathbf{v} \in Y$ . Nesse caso, dizemos que  $T$  é *invertível*.

DEFINIÇÃO 3.3. Uma transformação linear  $T : X \rightarrow Y$  entre dois espaços vetoriais sobre o mesmo corpo é dita **invertível** quando  $T$  é uma bijeção, i.e. é injetiva e sobrejetiva, o que significa dizer que  $T$  é um isomorfismo entre  $X$  e  $Y$ .

PROPOSIÇÃO 3.4. A inversa  $T^{-1} : Y \rightarrow X$  de uma transformação linear  $T : X \rightarrow Y$  entre dois espaços vetoriais sobre o mesmo corpo também é linear.

DEMONSTRAÇÃO. Deixamos essa demonstração com [Exercício 6.1](#).  $\square$

PROPOSIÇÃO 3.5. Se  $T : X \rightarrow Y$  e  $S : Y \rightarrow Z$  são transformações lineares ambas invertíveis, onde  $X, Y$  e  $Z$  são espaços vetoriais sobre o mesmo corpo, então

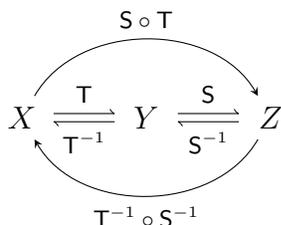


FIGURA 3.2. Ilustração da composição  $T \circ S : X \rightarrow Z$  de transformações  $T : X \rightarrow Y$  e  $S : Y \rightarrow Z$  e da inversa  $(T \circ S)^{-1} = S^{-1} \circ T^{-1} : Z \rightarrow X$ .

$(S \circ T) : X \rightarrow Z$  também é invertível e a sua inversa é dada pela composição de suas inversas, em ordem trocada, i.e.

$$(S \circ T)^{-1} = T^{-1} \circ S^{-1}.$$

DEMONSTRAÇÃO. Também deixamos essa demonstração como [Exercício 6.2](#).  $\square$

Em algumas situações, podemos ter uma inversa apenas em um sentido.

DEFINIÇÃO 3.4. Uma transformação linear  $T : X \rightarrow Y$  entre dois espaços vetoriais sobre o mesmo corpo é dita **invertível à direita** quando existe  $R : Y \rightarrow X$  linear tal que  $T \circ R = I$  e é dita **invertível à esquerda** quando existe  $L : Y \rightarrow X$  tal que  $L \circ T = I$ . Nesses casos,  $R$  é dita a **inversa à direita** de  $T$  e  $L$  é dita a **inversa à esquerda** de  $T$ . Naturalmente,  $T$  é invertível se, e somente se,  $T$  é invertível à direita e à esquerda e, nesse caso, as inversas coincidem. Além disso, se  $R$  é a inversa à direita de  $T$ , então  $T$  é a inversa à esquerda de  $R$  e se  $L$  é a inversa à esquerda de  $T$ , então  $T$  é a inversa à direita de  $L$ .

PROPOSIÇÃO 3.6. Seja  $T : X \rightarrow Y$  uma transformação linear entre dois espaços vetoriais  $X$  e  $Y$  sobre o mesmo corpo. Se  $T$  é invertível à direita, então  $T$  é sobrejetiva e se  $T$  é invertível à esquerda, então  $T$  é injetiva.

EXEMPLO 3.1. Considerando  $X = \mathbb{R}^2$  e  $Z = \mathbb{R}^1$ , vemos que  $T(x, y) = x$  é invertível à direita com inversa à direita  $S(z) = (z, 0)$ , pois  $T(S(z)) = T(z, 0) = z$ , mas  $T$  não é invertível à esquerda pois não é injetiva. Da mesma forma,  $S$  não é invertível à direita pois  $S$  não é sobrejetiva. Observe, ainda, que, nesse caso especial, podemos fazer a composição  $S \circ T$ , que nos dá  $S(T(x, y)) = S(x) = (x, 0)$ , mas, em geral, com três espaços distintos, apenas uma composição é possível. É interessante, também, enxergar as matrizes (nas bases canônicas) que representam cada transformação, com  $T$  sendo representado por  $\begin{bmatrix} 1 & 0 \end{bmatrix}$  e  $S$ , por  $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ .

### 3.3. Representação matricial em espaços de dimensão finita.

DEFINIÇÃO 3.5. *Dados espaços vetoriais  $X$  e  $Y$  de dimensão finita  $n = \dim(X) \in \mathbb{N}$  e  $m = \dim(Y) \in \mathbb{N}$  sobre um mesmo corpo e dadas bases  $\alpha = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  de  $X$  e  $\mathcal{b} = \{\mathbf{v}_1, \dots, \mathbf{v}_m\}$  de  $Y$ , a **representação matricial** de uma transformação linear  $T : X \rightarrow Y$  entre as bases é a matriz*

$$[T]_{\mathcal{b}}^{\alpha} = \begin{bmatrix} \langle \mathbf{g}_1, T(\mathbf{w}_1) \rangle & \langle \mathbf{g}_1, T(\mathbf{w}_2) \rangle & \cdots & \langle \mathbf{g}_1, T(\mathbf{w}_n) \rangle \\ \langle \mathbf{g}_2, T(\mathbf{w}_1) \rangle & \langle \mathbf{g}_2, T(\mathbf{w}_2) \rangle & \cdots & \langle \mathbf{g}_2, T(\mathbf{w}_n) \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{g}_m, T(\mathbf{w}_1) \rangle & \langle \mathbf{g}_m, T(\mathbf{w}_2) \rangle & \cdots & \langle \mathbf{g}_m, T(\mathbf{w}_n) \rangle \end{bmatrix}$$

onde  $\mathcal{b}^* = \{\mathbf{g}_1, \dots, \mathbf{g}_m\}$  é a base dual de  $\mathcal{b}$ . Quando  $X$  e  $Y$  são iguais e usamos a mesma base, digamos  $\mathcal{b}$ , nos dois espaços, escrevemos apenas  $[T]_{\mathcal{b}}$  para representar  $[T]_{\mathcal{b}}^{\mathcal{b}}$ .

TEOREMA 3.1. *Sejam  $X$  e  $Y$  dois espaços vetoriais sobre o mesmo corpo, com dimensões finitas  $n = \dim(X) \in \mathbb{N}$  e  $m = \dim(Y) \in \mathbb{N}$ . Sejam  $\alpha = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  e  $\mathcal{b} = \{\mathbf{v}_1, \dots, \mathbf{v}_m\}$  bases de  $X$  e  $Y$ , respectivamente. Então*

$$(T(\mathbf{u}))_{\mathcal{b}} = [T]_{\mathcal{b}}^{\alpha}(\mathbf{u})_{\alpha},$$

para todo  $\mathbf{u} \in X$ , onde  $(\mathbf{u})_{\alpha}$  é a representação de  $\mathbf{u}$  na base  $\alpha$ ;  $(T(\mathbf{u}))_{\mathcal{b}}$  é a representação de  $T\mathbf{u}$  na base  $\mathcal{b}$ ; e  $[T]_{\mathcal{b}}^{\alpha}$  é a representação de  $T$  da base  $\alpha$  para a base  $\mathcal{b}$ .

DEMONSTRAÇÃO. Observe que a representação de  $T(\mathbf{u})$  na base  $\mathcal{b}$  pode ser escrita como

$$(T(\mathbf{u}))_{\mathcal{b}} = (\langle \mathbf{g}_i, T(\mathbf{u}) \rangle)_{i=1, \dots, m},$$

para todo  $\mathbf{u} \in X$ . Para um tal  $\mathbf{u}$ , temos

$$\mathbf{u} = \langle \mathbf{f}_1, \mathbf{u} \rangle \mathbf{w}_1 + \cdots + \langle \mathbf{f}_n, \mathbf{u} \rangle \mathbf{w}_n,$$

de modo que, a  $i$ -ésima linha da representação de  $T(\mathbf{u})$  na base  $\mathcal{b}$  é

$$\begin{aligned} \langle \mathbf{g}_i, T(\mathbf{u}) \rangle &= \langle \mathbf{g}_i, T(\langle \mathbf{f}_1, \mathbf{u} \rangle \mathbf{w}_1 + \cdots + \langle \mathbf{f}_n, \mathbf{u} \rangle \mathbf{w}_n) \rangle \\ &= \langle \mathbf{g}_i, \langle \mathbf{f}_1, \mathbf{u} \rangle T(\mathbf{w}_1) + \cdots + \langle \mathbf{f}_n, \mathbf{u} \rangle T(\mathbf{w}_n) \rangle \\ &= \langle \mathbf{f}_1, \mathbf{u} \rangle \langle \mathbf{g}_i, T(\mathbf{w}_1) \rangle + \cdots + \langle \mathbf{f}_n, \mathbf{u} \rangle \langle \mathbf{g}_i, T(\mathbf{w}_n) \rangle \end{aligned}$$

Como isso vale para cada linha  $i$ , reconhecemos o vetor  $(\mathbb{T}(\mathbf{u}))_{\mathfrak{b}} = (\langle \mathfrak{g}_i, \mathbb{T}(\mathbf{u}) \rangle)_{i=1, \dots, m}$  como o produto matricial

$$\begin{bmatrix} \langle \mathfrak{g}_1, \mathbb{T}(\mathbf{w}_1) \rangle & \langle \mathfrak{g}_1, \mathbb{T}(\mathbf{w}_2) \rangle & \cdots & \langle \mathfrak{g}_1, \mathbb{T}(\mathbf{w}_n) \rangle \\ \langle \mathfrak{g}_2, \mathbb{T}(\mathbf{w}_1) \rangle & \langle \mathfrak{g}_2, \mathbb{T}(\mathbf{w}_2) \rangle & \cdots & \langle \mathfrak{g}_2, \mathbb{T}(\mathbf{w}_n) \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathfrak{g}_m, \mathbb{T}(\mathbf{w}_1) \rangle & \langle \mathfrak{g}_m, \mathbb{T}(\mathbf{w}_2) \rangle & \cdots & \langle \mathfrak{g}_m, \mathbb{T}(\mathbf{w}_n) \rangle \end{bmatrix} \begin{pmatrix} \langle \mathfrak{f}_1, \mathbf{u} \rangle \\ \langle \mathfrak{f}_2, \mathbf{u} \rangle \\ \vdots \\ \langle \mathfrak{f}_n, \mathbf{u} \rangle \end{pmatrix},$$

ou seja,

$$(\mathbb{T}(\mathbf{u}))_{\mathfrak{b}} = [\mathbb{T}]_{\mathfrak{b}}^{\mathfrak{a}}(\mathbf{u})_{\mathfrak{a}},$$

conforme desejado.  $\square$

Usando o **Teorema 2.1** do Núcleo e da Imagem, podemos dar uma representação simples de uma transformação qualquer em bases apropriadas.

**TEOREMA 3.2.** *Sejam  $X$  e  $Y$  dois espaços vetoriais sobre o mesmo corpo, com dimensões finitas  $n = \dim(X) \in \mathbb{N}$  e  $m = \dim(Y) \in \mathbb{N}$ , e seja  $\mathbb{T} : X \rightarrow Y$  uma transformação linear. Então existem bases  $\mathfrak{a}$  de  $X$  e  $\mathfrak{b}$  de  $Y$  tais que a representação  $[\mathbb{T}]_{\mathfrak{b}}^{\mathfrak{a}}$  de  $\mathbb{T}$  da base  $\mathfrak{a}$  para a base  $\mathfrak{b}$  tem a forma de blocos*

$$[\mathbb{T}]_{\mathfrak{b}}^{\mathfrak{a}} = \begin{bmatrix} \mathbf{I}_{n-k} & \mathbf{O}_{n-k,k} \\ \mathbf{O}_{m-n+k,n-k} & \mathbf{O}_{m-n+k,k} \end{bmatrix}.$$

**DEMONSTRAÇÃO.** Seja  $k = \dim(\ker(\mathbb{T}))$ . Se  $k = n$ , então  $\ker(\mathbb{T}) = X$  e a imagem é trivial, visto que  $\dim(\text{Im}(\mathbb{T})) = n - k = 0$ . Nesse caso,  $\mathbb{T}$  é o operador nulo,  $\mathbb{T} = \mathbf{O}$ , com

$$[\mathbb{T}]_{\mathfrak{b}}^{\mathfrak{a}} = \mathbf{O}_{m \times n},$$

em qualquer base, onde  $\mathbf{O}_{m \times n}$  é a matriz  $m \times n$  com todos os elementos nulos.

Suponha então  $k = \dim(\ker(\mathbb{T})) < n$ . Nesse caso, pelo **Teorema 2.1** do Núcleo e da Imagem,  $\dim(\text{Im}(\mathbb{T})) = n - k \geq 1$ . Obviamente,  $n - k \leq m$ , pois a dimensão da imagem não pode ser maior do que a do contra-domínio.

Tomemos uma base  $\{\mathbf{v}_1, \dots, \mathbf{v}_{n-k}\}$  da imagem  $\text{Im}(\mathbb{T})$ . Complete até uma base  $\mathfrak{b} = \{\mathbf{v}_1, \dots, \mathbf{v}_{n-k}, \mathbf{v}_{n-k+1}, \dots, \mathbf{v}_m\}$  de  $Y$ , se necessário (i.e. se  $n - k < m$ .)

Agora, para cada  $\mathbf{v}_i$ ,  $i = 1, \dots, n - k$ , que estão na imagem de  $\mathbb{T}$ , existe  $\mathbf{w}_i \in X$  tal que  $\mathbb{T}(\mathbf{w}_i) = \mathbf{v}_i$ .

Se  $k = \dim(\ker(\mathbb{T})) > 0$ , tome também uma base  $\{\mathbf{w}_{n-k+1}, \dots, \mathbf{w}_n\}$  do núcleo  $\ker(\mathbb{T})$ , formada por  $k$  vetores, numerados de  $n - k + 1$  até  $n$ . Se  $k = 0$ , então esse conjunto é vazio e podemos ignorar esses termos no argumento abaixo.

Vamos mostrar que  $\mathfrak{a} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  é uma base de  $X$ . Como são  $n$  vetores e  $\dim(X) = n$  (contabilidade já feita pelo **Teorema 2.1** do Núcleo e da Imagem), basta

verificar que são linearmente independentes. Suponha, então, que

$$\lambda_1 \mathbf{w}_1 + \cdots + \lambda_n \mathbf{w}_n = \mathbf{0},$$

onde  $\lambda_1, \dots, \lambda_n \in \mathbb{K}$ . Aplicando a transformação  $\mathbb{T}$  e usando que  $\mathbf{w}_{n-k+1}, \dots, \mathbf{w}_n$  estão no núcleo de  $\mathbb{T}$ , sobra

$$\lambda_1 \mathbb{T}(\mathbf{w}_1) + \cdots + \lambda_{n-k} \mathbb{T}(\mathbf{w}_{n-k}) = \mathbf{0}.$$

Como  $\mathbb{T}(\mathbf{w}_i) = \mathbf{v}_i$  e  $\{\mathbf{v}_1, \dots, \mathbf{v}_{n-k}\}$  é linearmente independente, então

$$\lambda_1 = \cdots = \lambda_{n-k} = 0.$$

Voltando a identidade inicial, sobra, agora,

$$\lambda_{n-k+1} \mathbf{w}_{n-k+1} + \cdots + \lambda_n \mathbf{w}_n = \mathbf{0}.$$

Como  $\{\mathbf{w}_{n-k+1}, \dots, \mathbf{w}_n\}$  é linearmente independente, por ser base de  $\ker(\mathbb{T})$ , então

$$\lambda_{n-k+1} = \cdots = \lambda_n = 0,$$

também. Isso completa a demonstração de que  $\mathcal{a} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  é base de  $X$ .

Agora falta escrever  $[\mathbb{T}]_{\mathcal{b}}^{\mathcal{a}}$ . Seja  $\mathcal{b}^* = \{\mathbf{g}_1, \dots, \mathbf{g}_m\}$  a base dual de  $\mathcal{b}$ . Cada coeficiente  $ij$  (linha  $i$ , coluna  $j$ ) da matriz  $[\mathbb{T}]_{\mathcal{b}}^{\mathcal{a}}$  é dado por

$$\langle \mathbf{g}_i, \mathbb{T}(\mathbf{w}_j) \rangle.$$

Para  $j = n - k + 1, \dots, n$ , temos  $\mathbf{w}_j \in \ker(\mathbb{T})$ , de modo que  $\langle \mathbf{g}_i, \mathbb{T}(\mathbf{w}_j) \rangle = 0$ , ou seja, todos os elementos das colunas  $n - k + 1$  até  $n$  se anulam.

Para  $i = n - k + 1, \dots, m$ , o funcional  $\mathbf{g}_i$  anula os elementos da imagem de  $\mathbb{T}$ , pois a imagem é gerado pelos vetores  $\mathbf{v}_1, \dots, \mathbf{v}_{n-k}$ . Portanto, todos os elementos das linhas  $n - k + 1$  até  $m$  se anulam.

Sobra o bloco  $i = 1, \dots, n - k$ ,  $j = 1, \dots, n - k$ . Nesse bloco,  $\mathbb{T}(\mathbf{w}_j) = \mathbf{v}_j$  e, por  $\mathcal{b}^*$  ser a base dual de  $\mathcal{b}$ , que é formada pelos  $\mathbf{v}_j$ 's, temos  $\langle \mathbf{g}_i, \mathbb{T}(\mathbf{w}_j) \rangle = \delta_{ij}$ , o Delta de Kronecker, nos dando a matriz identidade  $(n - k) \times (n - k)$ . Em resumo,

$$\langle \mathbf{g}_i, \mathbb{T}(\mathbf{w}_j) \rangle = \delta_{ij}, \quad i, j = 1, \dots, n - k,$$

e

$$\langle \mathbf{g}_i, \mathbb{T}(\mathbf{w}_j) \rangle = 0, \quad i = n - k + 1, \dots, m \text{ ou } j = n - k + 1, \dots, n.$$

Em outras palavras, podemos escrever, em forma de blocos,

$$[\mathbb{T}]_{\mathcal{b}}^{\mathcal{a}} = \begin{bmatrix} \mathbf{I}_{n-k} & \mathbf{O}_{n-k,k} \\ \mathbf{O}_{m-n+k,n-k} & \mathbf{O}_{m-n+k,k} \end{bmatrix}$$

conforme desejado. No caso particular em que  $k = 0$  e  $m > n$ , isso se reduz a

$$[\mathbb{T}]_{\mathcal{b}}^{\mathcal{a}} = \begin{bmatrix} \mathbf{I}_n \\ \mathbf{O}_{m-n,n} \end{bmatrix}.$$

Caso  $m = n - k$  (posto máximo), temos, por outro lado,

$$[\mathbb{T}]_{\mathcal{B}}^{\mathcal{A}} = [\mathbf{I}_{n-k} \quad \mathbf{O}_{n-k,k}].$$

E caso  $k = 0$  e  $m = n$ , temos apenas

$$[\mathbb{T}]_{\mathcal{B}}^{\mathcal{A}} = [\mathbf{I}_n].$$

□

OBSERVAÇÃO 3.4. Na demonstração do **Teorema 3.2**, trocando a ordem dos elementos da base  $\mathcal{A}$ , escolhendo primeiro uma base do núcleo  $\ker(\mathbb{T})$  e completando com as pré-imagens da base de  $\text{Im}(\mathbb{T})$ , a matriz muda de forma para

$$[\mathbb{T}]_{\mathcal{B}}^{\mathcal{A}} = \begin{bmatrix} \mathbf{O}_{n-k,k} & \mathbf{I}_{n-k} \\ \mathbf{O}_{m-n+k,k} & \mathbf{O}_{m-n+k,n-k} \end{bmatrix}.$$

Temos muita flexibilidade na representação, podendo escolher as bases  $\mathcal{A}$  e  $\mathcal{B}$  sem muita amarração.

EXEMPLO 3.2. Considere  $X = \mathbb{R}^2$ ,  $Y = \mathbb{R}^3$  e  $\mathbb{T}(x, y) = (x + y, x + y, 0)$ . O núcleo de  $\mathbb{T}$  tem dimensão 1 e é dado por  $\ker(\mathbb{T}) = \text{span}\{(1, -1)\}$ . A imagem de  $\mathbb{T}$  também tem dimensão 1 e é dada por  $\text{Im}(\mathbb{T}) = \text{span}\{(1, 1, 0)\}$ . Completamos até uma base  $\mathcal{B} = \{(1, 1, 0), (0, 1, 0), (0, 0, 1)\}$  de  $Y$ . Agora, uma pré-imagem de  $(1, 1, 0)$  é, por exemplo,  $(1, 0)$ , visto que  $\mathbb{T}(1, 0) = (1, 1, 0)$ , mas observe que qualquer elemento em  $(1, 0) + \ker(\mathbb{T}) = \{(1 + s, -s), s \in \mathbb{R}\}$  serve. Juntando a pré-imagem  $\{(1, 0)\}$  da base da imagem  $\text{Im}(\mathbb{T})$  com a base  $\{(1, -1)\}$  do núcleo  $\ker(\mathbb{T})$ , obtemos uma base  $\mathcal{B} = \{(1, 0), (1, -1)\}$  para  $X$ . As colunas de  $[\mathbb{T}]_{\mathcal{B}}^{\mathcal{A}}$  são obtidas vendo a ação de  $\mathbb{T}$  nos elementos da base de saída  $\mathcal{A}$  e representando-os na base de chegada  $\mathcal{B}$ . Observe que

$$\mathbb{T}(1, 0) = (1, 1, 0) = 1(1, 1, 0) + 0(0, 1, 0) + 0(0, 0, 1),$$

enquanto que

$$\mathbb{T}(1, 1) = (0, 0, 0) = 0(1, 1, 0) + 0(0, 1, 0) + 0(0, 0, 1).$$

Portanto,

$$[\mathbb{T}]_{\mathcal{B}}^{\mathcal{A}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{I}_1 & \mathbf{O}_{1 \times 1} \\ \mathbf{O}_{2 \times 1} & \mathbf{O}_{2 \times 1} \end{bmatrix}.$$

OBSERVAÇÃO 3.5. Não devemos nos iludir com a simplicidade da representação feita no **Teorema 3.2**. O que a transformação realmente faz está escondido na relação entre as bases. Outras representações são mais úteis, como a decomposição SVD, que veremos mais adiante. Assim como as representações de operadores lineares, feitas

usando-se uma mesma base de ida e de chegada, quando o domínio e o contradomínio coincidem.

EXEMPLO 3.3. Corroborando a ideia de que a representação acima esconde a ação da transformação na escolha das bases, considere uma rotação  $R_\theta : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  de um ângulo  $\theta$  no sentido trigonométrico, no plano, dada por

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

Na base canônica  $\mathfrak{b} = \{(1, 0), (0, 1)\}$ , temos a representação

$$[R_\theta]_{\mathfrak{b}} = \begin{bmatrix} \cos \theta & -\operatorname{sen} \theta \\ \operatorname{sen} \theta & \cos \theta \end{bmatrix}.$$

Mas, escolhendo a base  $\mathfrak{a} = \{(\cos \theta, -\operatorname{sen} \theta), (\operatorname{sen} \theta, \cos \theta)\}$ , temos

$$R_\theta(\cos \theta, -\operatorname{sen} \theta) = (\cos^2 \theta + \operatorname{sen}^2 \theta, \cos \theta \operatorname{sen} \theta - \operatorname{sen} \theta \cos \theta) = (1, 0)$$

e

$$R_\theta(\operatorname{sen} \theta, \cos \theta) = (\operatorname{sen} \theta \cos \theta - \cos \theta \operatorname{sen} \theta, \operatorname{sen}^2 \theta + \cos^2 \theta) = (0, 1),$$

de modo que

$$[R_\theta]_{\mathfrak{a}}^{\mathfrak{a}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

EXEMPLO 3.4. De maneira semelhante, considere a transformação  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  dada por  $T(x, y) = (2x, y/2)$ , para  $(x, y) \in \mathbb{R}^2$ . Na base canônica  $\mathfrak{e} = \{(1, 0), (0, 1)\}$ , essa transformação é representada por

$$[T]_{\mathfrak{e}} = \begin{bmatrix} 2 & 0 \\ 0 & 1/2 \end{bmatrix}.$$

Escolhendo, no entanto, a base  $\mathfrak{a} = \{(1/2, 0), (0, 2)\}$ , como  $T(1/2, 0) = (1, 0)$  e  $T(0, 2) = (0, 1)$ , temos a representação

$$[T]_{\mathfrak{a}}^{\mathfrak{a}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

#### 4. Transformação adjunta

Dados dois espaços vetoriais  $X$  e  $Y$  e considerando o produto de dualidade entre eles e os seus duais (ou quaisquer outras formas bilineares entre eles e outros espaços), há uma forma natural de gerar uma transformação “dual” de uma transformação linear de  $X$  em  $Y$ . A bilinearidade do produto de dualidade e a linearidade da transformação “primal” garantem a linearidade da transformação “dual”, chamada de *transformação adjunta*.

Isso é útil em vários contextos, onde um “problema primal” pode ser transformado em seu “problema dual” e, muitas vezes, ser resolvido mais facilmente assim. Veremos, posteriormente, que, quando os espaços  $X$  e  $Y$  são iguais e temos um produto interno disponível, a transformação adjunta pode ser representada como um operador no próprio espaço e, nesse caso, relações especiais entre o operador e seu adjunto podem ser exploradas.

#### 4.1. Adjunta de uma transformação.

DEFINIÇÃO 4.1. *Sejam  $X$  e  $Y$  espaços vetoriais sobre um mesmo corpo  $\mathbb{K}$  e seja  $T : X \rightarrow Y$  uma transformação linear. A **transformação adjunta**  $T^* : Y^* \rightarrow X^*$  é a transformação que leva um funcional linear  $\mathbf{g} \in Y^*$  em um funcional linear  $T^*\mathbf{g} \in X^*$  definido por sua ação  $(T^*\mathbf{g})(\mathbf{u}) = \langle T^*\mathbf{g}, \mathbf{u} \rangle_{X^*,X}$  em cada vetor  $\mathbf{u} \in X$  via*

$$\langle T^*\mathbf{g}, \mathbf{u} \rangle_{X^*,X} = \langle \mathbf{g}, T\mathbf{u} \rangle_{Y^*,Y}.$$

PROPOSIÇÃO 4.1. *A transformação adjunta  $T^* : Y^* \rightarrow X^*$  dada na [Definição 4.1](#) é linear.*

DEMONSTRAÇÃO. Basta verificar que

$$\begin{aligned} \langle T^*(\mathbf{g} + \gamma\mathbf{h}), \mathbf{u} \rangle_{Y^*,Y} &= \langle \mathbf{g} + \gamma\mathbf{h}, T\mathbf{u} \rangle_{X^*,X} = \langle \mathbf{g}, T\mathbf{u} \rangle_{X^*,X} + \gamma\langle \mathbf{h}, T\mathbf{u} \rangle_{X^*,X} \\ &= \langle T^*\mathbf{g}, \mathbf{u} \rangle_{Y^*,Y} + \gamma\langle T^*\mathbf{h}, \mathbf{u} \rangle_{Y^*,Y}, \quad \forall \mathbf{u} \in X, \end{aligned}$$

para todos  $\mathbf{g}, \mathbf{h} \in Y^*$  e  $\gamma \in \mathbb{K}$ . □

PROPOSIÇÃO 4.2. *Dados espaços vetoriais  $X, Y$  e  $Z$  sobre um mesmo corpo  $\mathbb{K}$ ,  $\lambda \in \mathbb{K}$  e transformações lineares  $T, L : X \rightarrow Y$ ,  $S : Y \rightarrow Z$ , então*

$$(T + L)^* = T^* + L^*, \quad (\lambda T)^* = \lambda T^*, \quad (S \circ T)^* = T^* \circ S^*.$$

*Além disso, caso  $T$  seja invertível, então*

$$(T^{-1})^* = (T^*)^{-1}.$$

OBSERVAÇÃO 4.1. A propriedade  $(\lambda T)^* = \lambda T^*$  mencionada na [Proposição 4.2](#) vale tanto em espaços vetoriais reais quanto em espaços vetoriais complexos. Mas ao considerarmos transformações entre espaços com produto interno e, junto com isso, identificarmos a transformação adjunta  $T^*$  com transformações  $\tilde{T}^*$  nos próprios espaços, é que temos uma diferenciação entre os casos real e complexo. No caso real, continuamos a ter a relação  $(\lambda\tilde{T})^* = \lambda\tilde{T}^*$ . Já no caso complexo, devido ao fato do produto interno complexo ser uma forma sesqui-linear (linear na primeira componente, linear-conjugada na segunda componente e conjugada-simétrica), a relação passa a ser  $(\lambda\tilde{T})^* = \bar{\lambda}\tilde{T}^*$ , onde  $\bar{\lambda}$  é o complexo conjugado de um escalar  $\lambda$ .

EXEMPLO 4.1 (Transformada de Fourier). Vimos no [Exemplo 1.4](#) que a transformada de Fourier pode ser considerada como um operador linear no espaço de Schwartz  $\mathcal{S}(\mathbb{R}^n, \mathbb{C})$ . O dual topológico do espaço de Schwartz pode ser visto como o *espaço das distribuições temperadas*  $\mathcal{S}(\mathbb{R}^n, \mathbb{C})'$ , quando o dual de  $L^2(\mathbb{R}^n)$  é identificado consigo mesmo. O espaço das distribuições temperadas está contido no *espaço das distribuições*  $\mathcal{D}(\mathbb{R}^n, \mathbb{R})'$ , que é o dual do espaço  $\mathcal{D}(\mathbb{R}^n, \mathbb{R})$  das funções teste infinitamente continuamente diferenciáveis de suporte compacto. Ou seja,

$$\mathcal{D}(\mathbb{R}^n, \mathbb{R}) \subset \mathcal{S}(\mathbb{R}^n, \mathbb{C}) \subset L^2(\mathbb{R}^n, \mathbb{C}) \equiv L^2(\mathbb{R}^n, \mathbb{C})' \subset \mathcal{S}(\mathbb{R}^n, \mathbb{C})' \subset \mathcal{D}(\mathbb{R}^n, \mathbb{R})'.$$

Usando essa dualidade, podemos definir a transformada (adjunta) de Fourier de distribuições temperadas  $F^* : \mathcal{S}(\mathbb{R}^n, \mathbb{C})' \rightarrow \mathcal{S}(\mathbb{R}^n, \mathbb{C})'$ . A transformada adjunta  $F^*$  coincide com transformada inversa  $G$  em  $L^2(\mathbb{R}^n, \mathbb{C})$ . Voltaremos a falar sobre isso quando discutirmos produto interno em espaços vetoriais complexos. De qualquer forma, a ideia é definir a transformada de Fourier de uma distribuição temperada via dualidade.

Em particular, podemos falar da transformada de Fourier da distribuição delta de Dirac  $\delta$  e observar que  $\hat{\delta}(\boldsymbol{\omega}) = (2\pi)^{-n/2}$ ,  $\forall \boldsymbol{\omega} \in \mathbb{R}^n$ , ou seja, o seu espectro contém todas as frequências, com igual amplitude!

De fato,

$$\langle \mathcal{F}\delta, \mathbf{u} \rangle = \langle \delta, \mathcal{F}^*\mathbf{u} \rangle = \langle \delta, \mathcal{G}\mathbf{u} \rangle = (\mathcal{G}\mathbf{u})(\mathbf{0}) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \mathbf{u}(\mathbf{x}) \, dx = \left\langle \frac{1}{(2\pi)^{n/2}}, \mathbf{u} \right\rangle,$$

**4.2. Representação matricial de transformações adjuntas.** Ao olharmos a representação matricial da adjunta  $T^*$  de uma transformação linear  $T$  entre espaços de dimensão finita, observamos que a representação da adjunta é a transposta da matriz que representa a transformação  $T$ . Lembremos que a matriz transposta é obtida dispendo as linhas de uma matriz em colunas (veja [Exemplo 1.23](#)).

TEOREMA 4.1. *Sejam  $X$  e  $Y$  espaços vetoriais de dimensão finita  $n = \dim(X) \in \mathbb{N}$  e  $m = \dim(Y) \in \mathbb{N}$ . Seja  $T \in \mathcal{L}(X, Y)$  uma transformação linear de  $X$  em  $Y$ . Dadas bases  $\alpha$ , de  $X$ , e  $\beta$ , de  $Y$ , sejam  $\alpha'$  e  $\beta'$  as respectivas bases duais. A representação matricial  $[T^*]_{\alpha'}^{\beta'}$ , da transformação adjunta é a transposta da representação matricial  $[T]_{\beta}^{\alpha}$ .*

DEMONSTRAÇÃO. Sejam  $\alpha = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ ,  $\beta = \{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ ,  $\alpha' = \{\mathbf{f}_1, \dots, \mathbf{f}_n\}$  e  $\beta' = \{\mathbf{g}_1, \dots, \mathbf{g}_m\}$ .

De acordo com a [Definição 3.5](#), temos

$$[T]_{\beta}^{\alpha} = (a_{ij})_{i,j=1}^{i=m, j=n}, \quad a_{ij} = \langle \mathbf{g}_i, T(\mathbf{w}_j) \rangle_{Y^*, Y}$$

Para a representação da transformação adjunta, lembremos, do **Teorema 4.1**, que o bidual  $X^{**}$  é isomorfo a  $X$  via isomorfismo canônico  $J : X \rightarrow X^{**}$ , dado por

$$\langle J(\mathbf{u}), \mathbf{f} \rangle_{X^{**}, X^*} = \langle \mathbf{f}, \mathbf{u} \rangle_{X^*, X}.$$

Com isso, a base  $\mathcal{b}''$  de  $X^{**}$  dual a  $\mathcal{b}'$  está associada exatamente a  $\mathcal{b}$ , via  $\mathcal{b}'' = \{J(\mathbf{w}_1), \dots, J(\mathbf{w}_n)\}$ .

Dessa forma, temos a representação  $[\mathbf{T}^*]_{\mathcal{a}'}$  dada por

$$[\mathbf{T}^*]_{\mathcal{a}'}^{\mathcal{b}'} = (b_{ij})_{i=1, \dots, n, j=1, \dots, m}^{\mathcal{b}'}, \quad b_{ij} = \langle J(\mathbf{w}_i), \mathbf{T}^*(\mathbf{g}_j) \rangle_{X^{**}, X^*} = \langle \mathbf{T}^*(\mathbf{g}_j), \mathbf{w}_i \rangle_{X^*, X}.$$

Agora, usando a definição de transformação adjunta, temos

$$b_{ij} = \langle \mathbf{T}^*(\mathbf{g}_j), \mathbf{w}_i \rangle_{X^*, X} = \langle \mathbf{g}_j, \mathbf{T}\mathbf{w}_i \rangle_{Y^*, Y} = a_{ji},$$

para  $i = 1, \dots, n$  e  $j = 1, \dots, m$ . Isso mostra que  $[\mathbf{T}^*]_{\mathcal{a}'}^{\mathcal{b}'}$  é a matriz transposta de  $[\mathbf{T}]_{\mathcal{b}}^{\mathcal{a}}$ .  $\square$

**OBSERVAÇÃO 4.2.** A relação entre as representações de  $\mathbf{T}$  e  $\mathbf{T}^*$ , de uma ser a transposta da outra, está atrelada à escolha das bases de  $X^*$  e  $Y^*$  serem as bases duais às escolhidas para  $X$  e  $Y$ . Isso é importante ressaltar.

**OBSERVAÇÃO 4.3.** Tendo em vista a **Observação 4.2**, vale destacar o que acontece no caso em que tratamos diretamente com  $\mathbb{R}^n$  e com matrizes, ao invés de partirmos de uma transformação e escolhermos uma base para representá-la. Nesse caso, partindo de uma matriz, é como se tivéssemos, implicitamente, escolhido as bases canônicas. Ainda assim, a adjunta não é, naturalmente, uma matriz. Mas veremos, quando falarmos de produto interno, no **Capítulo 7**, que, ao considerarmos o produto escalar em  $\mathbb{R}^n$ , podemos associar a adjunta da transformação linear associada a uma matriz  $\mathbf{A} \in \mathbb{R}^{n \times n}$  à transformação linear associada à transposta  $\mathbf{A}^{\text{tr}}$  de  $\mathbf{A}$ . No caso de uma matriz complexa  $\mathbf{A} \in \mathbb{C}^{n \times n}$  em  $\mathbb{C}^n$ , a adjunto estará associada à matriz adjunta, que é o complexo conjugado da matriz transposta.

**EXEMPLO 4.2.** Considere a transformação linear  $\mathbf{T}(x, y, z) = (x + y, y + z)$  de  $\mathbb{R}^3$  em  $\mathbb{R}^2$ . Escolhendo as bases canônicas  $\mathbf{e}_3$  de  $\mathbb{R}^3$  e  $\mathbf{e}_2$  de  $\mathbb{R}^2$ , temos a representação

$$[\mathbf{T}]_{\mathbf{e}_2}^{\mathbf{e}_3} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

As bases duais  $\mathbf{e}'_n$  são transformações que capturam cada coordenada dos vetores, i.e.  $\mathbf{f}_j(x_1, \dots, x_n) = x_j$ ,  $j = 1, \dots, n$ ,  $n \in \mathbb{N}$ . Assim,

$$[\mathbf{T}^*]_{\mathbf{e}'_3}^{\mathbf{e}'_2} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

### 4.3. Relações entre imagem e núcleo de uma transformação e de sua adjunta.

TEOREMA 4.2. *Dada uma transformação linear  $T \in \mathcal{L}(X, Y)$  entre espaços vetoriais  $X, Y$  sobre um mesmo corpo  $\mathbb{K}$ , temos  $\text{Im}(T)^0 = \ker(T^*)$ , ou seja, o anulador da imagem de  $T$  é exatamente o núcleo da adjunta de  $T$ .*

DEMONSTRAÇÃO. Pelo produto de dualidade e a definição de adjunta, temos

$$\langle T^* \mathbf{g}, \mathbf{u} \rangle = \langle \mathbf{g}, T\mathbf{u} \rangle,$$

para todo  $\mathbf{u} \in X$  e  $\mathbf{g} \in Y^*$ . Logo, usando o [Corolário 2.3](#),

$$\begin{aligned} \mathbf{g} \in \ker(T^*) &\Leftrightarrow T^* \mathbf{g} = \mathbf{o} \Leftrightarrow \langle T^* \mathbf{g}, \mathbf{u} \rangle = 0, \forall \mathbf{u} \in X \Leftrightarrow \langle \mathbf{g}, T\mathbf{u} \rangle = 0, \forall \mathbf{u} \in X \\ &\Leftrightarrow \langle \mathbf{g}, \mathbf{v} \rangle = 0, \forall \mathbf{v} \in \text{Im}(T) \Leftrightarrow \mathbf{g} \in \text{Im}(T)^0, \end{aligned}$$

provando o resultado.  $\square$

TEOREMA 4.3. *Dada uma transformação linear  $T \in \mathcal{L}(X, Y)$  entre espaços vetoriais  $X, Y$  sobre um mesmo corpo  $\mathbb{K}$ , temos  $\ker(T^*)^0 = \text{Im}(T)$ , ou seja, o anulador do núcleo da adjunta  $T^*$  é exatamente a imagem de  $T$ .*

DEMONSTRAÇÃO. Usando que o anulador do anulador de um subespaço é o próprio subespaço (veja [Exercício 7.12](#)), obtemos, a partir do resultado do [Teorema 4.2](#), que

$$\text{Im}(T) = \text{Im}(T)^{00} = \ker(T^*)^0,$$

provando o resultado.  $\square$

TEOREMA 4.4. *Dada uma transformação linear  $T \in \mathcal{L}(X, Y)$  entre espaços vetoriais  $X, Y$  sobre um mesmo corpo  $\mathbb{K}$  e supondo que  $X$  tem dimensão finita, então*

$$\dim \text{Im}(T) = \dim \text{Im}(T^*),$$

*ou seja, o posto da adjunta é igual ao posto da transformação.*

DEMONSTRAÇÃO. Usando a identidade dada no [Teorema 5.1](#), aplicada a  $S = \text{Im}(T)$ , temos

$$\dim X = \dim \text{Im}(T) + \dim \text{Im}(T^0).$$

Por outro lado, usando o [Teorema 2.1](#) do Núcleo e da Imagem no espaço dual, temos

$$\dim X^* = \dim \ker(T^*) + \dim \text{Im}(T^*).$$

Como o dual tem a mesma dimensão do espaço, obtemos

$$\dim \text{Im}(T) + \dim \text{Im}(T^0) = \dim \ker(T^*) + \dim \text{Im}(T^*).$$

Pelo **Teorema 4.2**, temos  $\dim(\text{Im}(\mathbb{T})^0) = \dim(\ker(\mathbb{T}^*))$ , logo, obtemos a identidade

$$\dim \text{Im}(\mathbb{T}) = \dim \text{Im}(\mathbb{T}^*),$$

como queríamos demonstrar.  $\square$

**OBSERVAÇÃO 4.4.** O resultado do **Teorema 4.4** se traduz para matrizes como o resultado clássico, em Álgebra Linear, de que “posto-linha” é igual ao “posto-coluna”.

Caso o contradomínio  $Y$  tenha a mesma dimensão que o domínio (em particular caso  $\mathbb{T}$  seja um operador linear), então obtemos também uma igualdade entre as dimensões dos núcleos.

**TEOREMA 4.5.** *Seja  $\mathbb{T} \in \mathcal{L}(X, Y)$  uma transformação linear entre espaços vetoriais  $X, Y$  de dimensão finita, sobre um mesmo corpo  $\mathbb{K}$ , com  $\dim X = \dim Y$ . Então  $\dim \ker(\mathbb{T}) = \dim \ker(\mathbb{T}^*)$ .*

**DEMONSTRAÇÃO.** Pelo **Teorema 2.1** do Núcleo e da Imagem, temos

$$\dim(X) = \dim \ker \mathbb{T} + \dim \text{Im } \mathbb{T}, \quad \dim(Y^*) = \dim \ker \mathbb{T}^* + \dim \text{Im } \mathbb{T}^*.$$

Como  $\dim(X) = \dim(Y)$  e como  $\dim(Y) = \dim(Y^*)$ , obtemos

$$\dim \ker \mathbb{T} + \dim \text{Im } \mathbb{T} = \dim \ker \mathbb{T}^* + \dim \text{Im } \mathbb{T}^*.$$

Do **Teorema 4.4**, as dimensões das imagens coincidem, nos dando

$$\dim \ker \mathbb{T} = \dim \ker \mathbb{T}^*,$$

como desejado.  $\square$

**EXEMPLO 4.3.** Consideremos, novamente, a transformação do **Exemplo 4.2**,  $\mathbb{T}(x, y, z) = (x + y, y + z)$  de  $\mathbb{R}^3$  em  $\mathbb{R}^2$ , representada na base canônica por

$$[\mathbb{T}]_{e_2}^{e_3} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

e cuja adjunta é representada por

$$[\mathbb{T}^*]_{e'_3}^{e'_2} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Temos

$$\ker(\mathbb{T}) = \{(x, y, z); y = -x, z = x\} = \text{span}\{(1, -1, 1)\}, \quad \text{Im}(\mathbb{T}) = \mathbb{R}^2,$$

enquanto que ...

## 5. Aplicações

**5.1. Sistemas de equações lineares, forma escalonada e decomposição LU.** A representação em blocos dada no Teorema 3.2 é quase a forma escalonada reduzida vista nos cursos iniciais de Álgebra Linear, que visa a resolução de sistemas de equações algébricas lineares. A diferença é que, na forma escalonada, não alteramos a base de saída  $\mathcal{a}$ , como feito no Teorema 3.2. Apenas alteramos a base  $\mathcal{b}$ .

O interesse inicial é na resolução de sistemas de equações algébricas lineares, da forma

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2, \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m, \end{cases} \quad (5.1)$$

onde  $\mathbf{A} = (a_{ij})_{i,j=1}^{i=m,j=n}$  e  $\mathbf{b} = (b_i)_{i=1}^m$  são conhecidos e  $\mathbf{u} = (x_j)_{j=1}^n$  é formado pelas variáveis que queremos encontrar e que resolvem o sistema. Isso pode ser escrito na forma vetorial

$$\mathbf{A}\mathbf{u} = \mathbf{b}.$$

Denotando a base canônica no domínio  $\mathbb{R}^n$  por  $\mathbf{e}_n = \{\mathbf{e}_1^n, \dots, \mathbf{e}_n^n\}$  e base canônica no contradomínio  $\mathbb{R}^m$  por  $\mathbf{e}_m = \{\mathbf{e}_1^m, \dots, \mathbf{e}_m^m\}$ , a matriz  $\mathbf{A}$  está associada a uma transformação linear  $\mathbb{T}_{\mathbf{A}} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  com

$$[\mathbb{T}_{\mathbf{A}}]_{\mathbf{e}_m}^{\mathbf{e}_n} = \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

O método de eliminação gaussiana consiste em fazer operações *elementares* nas linhas do sistema até chegar a uma forma triangular superior para um sistema equivalente ao original. O que essas operações nas linhas fazem é, no fundo, escolher uma nova base  $\mathcal{b}$  para o espaço de chegada. Mas o espaço de saída continua com a base canônica  $\mathbf{e}_n$ , por isso as variáveis  $x_1, \dots, x_n$  para as soluções continuam as mesmas. Nesse processo, podemos chegar na forma escalonada ou na forma escalonada reduzida.

Uma maneira de enxergar a forma canônica reduzida (*reduced row-echelon form*) é escolher o primeiro elemento da base  $\mathcal{b}$  como sendo a primeira coluna não nula da matriz  $\mathbf{A}$ ,

$$\mathbf{v}_1 = \mathbf{A}\mathbf{e}_{j_1}^n \neq \mathbf{0},$$

com  $\mathbf{A}\mathbf{e}_j = \mathbf{0}$  para  $1 \leq j \leq j_1$ , se houver. Em seguida, buscamos a primeira coluna  $j_2$ ,  $j_1 < j_2 \leq n$  tal que

$$\mathbf{v}_2 = \mathbf{A}\mathbf{e}_{j_2}^n \notin \text{span}\{\mathbf{v}_1\},$$

e assim por diante, encontrando uma sequência de colunas  $j = j_1, j_2, \dots, j_k$  tal que

$$\mathbf{v}_l = \mathbf{A}\mathbf{e}_{j_l}^n \notin \text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_{l-1}\},$$

com

$$\mathbf{A}\mathbf{e}_j^n \in \text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_{l-1}\},$$

para  $j_{l-1} \leq j < j_l$ . Caso  $k < m$ , simplesmente completamos com elementos quaisquer até uma base  $\mathcal{b} = \{\mathbf{v}_1, \dots, \mathbf{v}_m\}$  de  $\mathbb{R}^m$ . Assim, da base  $\mathcal{e}_n$  à base  $\mathcal{b}$ , obtemos a forma canônica reduzida

$$[\mathbf{T}]_{\mathcal{b}}^{\mathcal{e}_n} = \begin{bmatrix} 1 & \star & \star & 0 & 0 & \star & 0 & \star & 0 \\ 0 & 0 & 1 & \star & 0 & \star & 0 & \star & 0 \\ 0 & 0 & 0 & 1 & \star & \star & 0 & \star & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \star & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Como a base de saída não mudou, as soluções  $(x_1, \dots, x_n)$  do sistema original são as mesmas do sistema equivalente na forma reduzida,

$$\begin{bmatrix} 1 & \star & \star & 0 & 0 & \star & 0 & \star & 0 \\ 0 & 0 & 1 & \star & 0 & \star & 0 & \star & 0 \\ 0 & 0 & 0 & 1 & \star & \star & 0 & \star & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \star & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \tilde{b}_1 \\ \tilde{b}_2 \\ \vdots \\ \tilde{b}_n \end{pmatrix},$$

para um lado direito apropriado  $\tilde{\mathbf{b}} = (\tilde{b}_1, \dots, \tilde{b}_n)$ .

A única dificuldade, aqui, é encontrar  $\tilde{\mathbf{b}}$ , que é a representação de  $\mathbf{b}$  na nova base  $\mathcal{b}$ . Encontrar  $\tilde{\mathbf{b}}$  requer inverter o processo de mudança de base. É nesse ponto que entra o método de eliminação gaussiana. O método de eliminação gaussiana é mais do que meramente encontrar uma nova base tornando o sistema triangular. É uma receita para se calcular a mudança inversa de base e, em particular, o novo lado direito  $\tilde{\mathbf{b}}$ , facilitando os cálculos. O ingrediente principal dessa receita são as operações elementares. As operações elementares são invertíveis e as suas inversas também são operações elementares. Vamos aos detalhes.

As operações elementares são as de

- (i) Multiplicação de uma linha por um escalar não nulo, que tem a forma matricial seguinte, com inversa da mesma forma,

$$\mathbf{E} = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & \lambda & & \\ & & & & 1 & \\ & & & & & \ddots \\ & & & & & & 1 \end{bmatrix}, \quad \mathbf{E}^{-1} = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & \frac{1}{\lambda} & & \\ & & & & 1 & \\ & & & & & \ddots \\ & & & & & & 1 \end{bmatrix},$$

;

- (ii) Adição de um múltiplo de uma linha a uma outra linha, que tem a seguinte forma matricial, com inversa da mesma forma,

$$\mathbf{E} = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & \ddots & & \\ & & \lambda & & 1 & \\ & & & & & \ddots \\ & & & & & & 1 \end{bmatrix}, \quad \mathbf{E}^{-1} = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & \ddots & & \\ & & & & 1 & \\ & & & & & \ddots \\ & & & & & & 1 \end{bmatrix}$$

- (iii) Troca posição de duas linhas, que tem a seguinte forma matricial, com mesma inversa

$$\mathbf{E} = \mathbf{E}^{-1} = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 0 & & 1 & \\ & & & \ddots & & \\ & & 1 & & 0 & \\ & & & & & \ddots \\ & & & & & & 1 \end{bmatrix}$$

Para a eliminação gaussiana, a ideia também começa com a escolha da primeira coluna  $j_1$  não nula da matriz  $\mathbf{A}$ ,

$$\mathbf{A}\mathbf{e}_{j_1}^n \neq \mathbf{0}$$

e nos dando o primeiro elemento da nova base  $\mathcal{b}$ ,

$$\mathbf{v}_1 = \mathbf{A}\mathbf{e}_{j_1}^n,$$

mas a diferença importante é a de nos dar uma receita para escrever  $\mathbf{e}_1^m$ , que é a representação de  $\mathbf{v}_1$  na base desejada  $\mathcal{b}$ , em termos de  $\mathbf{v}_1$ . Essa receita vem na forma de uma composição  $\mathbf{E}_1 = \mathbf{E}_{1,r_1} \mathbf{E}_{1,r_1-1} \cdots \mathbf{E}_{1,1}$  de  $r_1 \in \mathbb{N}$  operações elementares  $\mathbf{E}_{1,r}$ ,  $r = 1, \dots, r_1$ , i.e.

$$[\mathbf{v}_1]_{\mathcal{b}} = \mathbf{e}_1^m = \mathbf{E}_1 \mathbf{v}_1 = \mathbf{E}_1 \mathbf{A} \mathbf{e}_{j_1}^n.$$

O novo sistema, em particular o novo lado direito do sistema, com esse primeiro escalonamento, toma a forma

$$\mathbf{E}_1 \mathbf{A} \mathbf{u} = \mathbf{E}_1 \mathbf{b}.$$

Se  $j_1 = 1$ , a primeira coluna de  $[\mathbf{T}]_{\mathcal{b}}^{\mathcal{e}_n}$  toma a forma

$$\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

e a matriz escalonada toma a forma

$$\mathbf{E}_1 \mathbf{A} = \begin{bmatrix} 1 & \star & \cdots & \star \\ 0 & \star & \cdots & \star \\ \vdots & \vdots & \cdots & \vdots \\ 0 & \star & \cdots & \star \end{bmatrix}.$$

mas, caso  $j_1 > 1$ , então as primeiras  $j_1 - 1$  colunas são nulas e a matriz escalonada toma a forma

$$\mathbf{E}_1 \mathbf{A} = \begin{bmatrix} 0 & \cdots & 0 & 1 & \star & \cdots & \star \\ 0 & \cdots & 0 & 0 & \star & \cdots & \star \\ \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & 0 & 0 & \star & \cdots & \star \end{bmatrix}.$$

Em seguida, pegamos o menor índice  $j_2$ ,  $j_1 < j_2 \leq n$ , tal que

$$\mathbf{A} \mathbf{e}_{j_2}^n \notin \text{span}\{\mathbf{v}_1\},$$

o que é equivalente a

$$\mathbf{E}_1 \mathbf{A} \mathbf{e}_{j_2}^n \notin \text{span}\{\mathbf{E}_1 \mathbf{v}_1\} = \text{span}\{\mathbf{e}_1^m\},$$

Veja que, com a matriz escalonada, fica fácil decidir se  $\mathbf{E}_1 \mathbf{A} \mathbf{e}_{j_2}^n$  está no conjunto gerado por  $\mathbf{e}_1^m$  ou não, ou seja, basta olhar se há algum coeficiente não nulo do segundo coeficiente em diante.

A outra diferença, agora, é que o segundo vetor  $\mathbf{v}_2$  da base  $\mathcal{b}$  é tomado de tal forma que

$$\mathbf{E}_2 \mathbf{E}_1 \mathbf{e}_{j_2}^n = a_{12}^1 \mathbf{e}_1^m + \mathbf{e}_2^m,$$

para um escalar  $a_{12}^1$  apropriado, dado pela primeira coordenada de  $\mathbf{E}_1 \tilde{\mathbf{e}}_2^n$ , e com as operações elementares compondo  $\mathbf{E}_2$  envolvendo apenas a segunda linha em diante.

Seguimos assim, até encontrar um conjunto  $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$  linearmente independente, para algum  $k \leq \min\{n, m\}$ , e uma sequência  $\mathbf{E}_1, \dots, \mathbf{E}_k$  de composições de operações elementares, tais que o sistema se escreve de forma equivalente como

$$\mathbf{E}_k \cdots \mathbf{E}_1 \mathbf{A} \mathbf{u} = \mathbf{E}_k \cdots \mathbf{E}_1 \mathbf{b},$$

com  $\mathbf{U} = \mathbf{E}_k \cdots \mathbf{E}_1 \mathbf{A}$  na forma escada, com pivôs iguais a 1. Essa forma escada é, por exemplo, da forma

$$\mathbf{U} = \mathbf{E}_k \cdots \mathbf{E}_1 \mathbf{A} = \begin{bmatrix} 1 & * & * & * & * & * & * & * & * \\ 0 & 0 & 1 & * & * & * & * & * & * \\ 0 & 0 & 0 & 1 & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Em certos casos, as matrizes elementares não necessitam da operação de troca de linhas. Nesse caso, todas as matrizes elementares são triangulares inferiores, assim como as suas inversas, de modo que

$$\mathbf{L} = \mathbf{E}_1^{-1} \mathbf{E}_2^{-1} \cdots \mathbf{E}_k^{-1}. \quad (5.2)$$

é triangular inferior e obtemos  $\mathbf{A}$  como um produto de uma matriz triangular inferior  $\mathbf{L}$  com uma matriz triangular superior  $\mathbf{U}$ , conhecida como **decomposição LU** de  $\mathbf{A}$ :

$$\mathbf{A} = \mathbf{L} \mathbf{U}. \quad (5.3)$$

Caso a troca de linhas seja necessária, a matriz  $\mathbf{L}$  não é mais triangular inferior e estamos efetivamente falando do processo de *pivoteamento*, que implica em uma permutação das linhas. Mais detalhes sobre isso em uma disciplina de Álgebra Linear Computacional.

**5.2. Diferenciação automática.** Diferenciação é um dos pilares da matemática e é fundamental em diversas aplicações, não apenas para a formulação de modelos, mas também para a estimação de parâmetros, análise de sensibilidade, inferência, etc.. Para efeitos práticos, dada a complexidade de diversos modelos, é importante termos à disposição ferramentas computacionais de diferenciação. Uma ferramenta que tem se mostrado extremamente eficiente é a de diferenciação automática. Ela é baseada em reduzir uma expressão complexa a uma composição de funções primitivas, para as quais a diferencial é conhecida e, com isso, obter a diferencial da função composta de maneira iterativa, a partir da composição dessas primitivas.

A diferenciação é, na verdade, um problema muito mais de Análise do que Álgebra Linear, mas, no final das contas, envolve a composição de transformações lineares. Além disso, para ser eficiente, ela deve ter o maior número possível de primitivas e isso envolve obter primitivas de operações envolvendo, também, diversas transformações lineares (e.g. traço, determinante e inversa de um operador, assim como o de diversas fatorizações de uma transformação linear, entre outras), que, no fundo, são o pilar da computação científica. Por esses motivos, é interessante tomarmos um certo espaço, aqui, para descrever essa ferramenta.

Para colocar o problema em perspectiva, um caminho natural para a diferenciação é a *diferenciação simbólica*, onde se busca expressar explicitamente uma determinada composição de funções em termos da variável inicial.

Por exemplo, dadas fórmulas de funções  $f = f(x)$  e  $g = g(y)$ , com derivadas  $f'(x)$  e  $g'(y)$ , buscamos uma fórmula explícita, em função de  $x$ , para a derivada  $h'(x)$  da composição  $h(x) = g(f(x))$ . Para efeito de ilustração, se  $f(x) = \cos(x)$  e  $g(y) = x^2$ , então  $h(x) = g(f(x)) = \cos(x)^2$  e  $h'(x) = -2 \cos(x) \sin(x)$ . Isso, no entanto, pode se tornar bastante custoso ao considerarmos composições um pouco mais complexas, como

$$h(x) = \sin(2x + 2e^{2\cos(x)} + \ln(1 + x^3)). \quad (5.4)$$

Nesse caso, a derivada se torna

$$h'(x) = \cos(2x + 2e^{2\cos(x)} + \ln(1 + x^3)) \left( 2 - 4 \sin(x) e^{2\cos(x)} + \frac{3x^2}{1 + x^3} \right).$$

A complexidade da derivada aumenta em uma ordem considerável com a complexidade da composição. Funções de várias variáveis complicam ainda mais.

Aliás, vale ressaltar que, mesmo no caso de uma única variável, como a  $h(x)$  em (5.4), temos tanto operadores unários (i.e. que dependem apenas de uma variável, como  $x \mapsto 2x$ ,  $x \mapsto x^3$ ,  $x \mapsto \cos(x)$ ,  $x \mapsto e^x$ ,  $x \mapsto \ln(x)$  etc.) quanto binários (como  $(x, y) \mapsto x + y$  e  $(x, y) \mapsto xy$ ), sendo que os binários são associativos e podem envolver mais de dois argumentos (como em  $(x, y, z) \mapsto x + y + z$ , e assim por diante) de modo que, invariavelmente, mesmo no caso escalar, devemos tratar funções multivariadas em algum momento.

Podemos usar as ferramentas de introspeção da linguagem julia para ver como a expressão acima é entendida, em uma primeira etapa, por uma linguagem de programação. Observe, a seguir, que as operações binárias aparecem como diferentes ramos (que podem ter mais de dois, no caso de múltiplas operações binárias associativas).

```
julia> dump(Meta.parse("(sin(2x + 2exp(cos(x)) + log(1 + x^3))")); maxdepth=20)
Expr
```

```

head: Symbol quote
args: Array{Any}((1,))
1: Expr
  head: Symbol call
  args: Array{Any}((2,))
  1: Symbol sin
  2: Expr
    head: Symbol call
    args: Array{Any}((4,))
    1: Symbol +
    2: Expr
      head: Symbol call
      args: Array{Any}((3,))
      1: Symbol *
      2: Int64 2
      3: Symbol x
    3: Expr
      head: Symbol call
      args: Array{Any}((3,))
      1: Symbol *
      2: Int64 2
      3: Expr
        head: Symbol call
        args: Array{Any}((2,))
        1: Symbol exp
        2: Expr
          head: Symbol call
          args: Array{Any}((2,))
          1: Symbol cos
          2: Symbol x
    4: Expr
      head: Symbol call
      args: Array{Any}((2,))
      1: Symbol log
      2: Expr
        head: Symbol call
        args: Array{Any}((3,))
        1: Symbol +
        2: Int64 1
        3: Expr
          head: Symbol call
          args: Array{Any}((3,))
          1: Symbol ^
          2: Symbol x
          3: Int64 3

```

Isso pode ser visualizado em forma de árvore, como ilustrado na [Figura 5.1](#).

O que é evidente é a complexidade da expressão e o custo de se obter a derivada simbólica da mesma. A derivação simbólica é possível e é muito útil. Há pacotes que

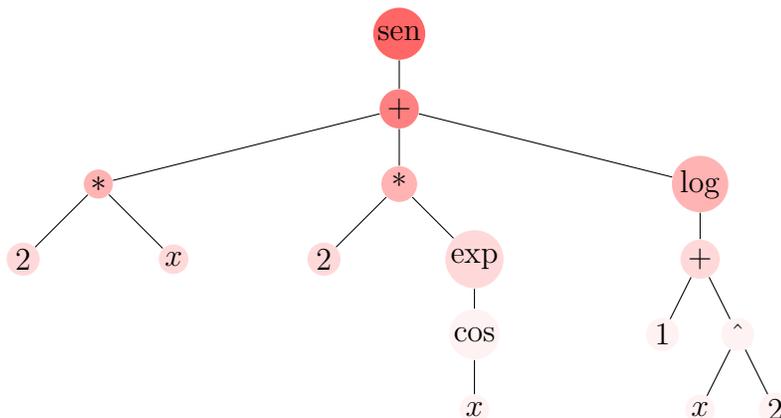


FIGURA 5.1. Visualização, em forma de árvore, de como a linguagem interpreta e expressa, inicialmente, a função  $h(x) = \text{sen}(2x + 2e^{2 \cos(x)} + \ln(1 + x^3))$ .

se utilizam de representações simbólicas de um modelo para construir automaticamente uma implementação direta do modelo, inclusive de forma otimizada (veja, por exemplo, [ModelingToolkit.jl: High-Performance Symbolic-Numeric Equation-Based Modeling](#)) [10].

Em muitos casos, no entanto, como por exemplo em redes neurais, em modelos envolvendo equações diferenciais, em modelos probabilísticos, e assim por diante, a derivação simbólica não é uma opção viável.

Uma alternativa é aproximar a derivada via diferenças finitas. Por um lado, esse método é, em princípio, bem menos custoso e de mais simples e ampla implementação, mas está sujeito a erros não desprezíveis de aproximação. Mesmo assim, pode ser bastante conveniente.

Um terceira alternativa é a de *derivação automática*, onde as derivadas de uma ampla classe de funções fundamentais são implementadas de maneira exata (apenas sujeitas a erros de arredondamento de ponto flutuante mas sem erros mais grosseiros de aproximação) mas não se busca expressar as derivadas da função composta explicitamente. Essas são calculadas recursivamente, somente nos pontos em que isso se faz necessário.

Relacionadas ao exemplo acima, podemos implementar as derivadas parciais das seguintes funções

$f(x)$	$f'(x)$
$c$	$0$
$x$	$1$
$x^2$	$2x$
$x^3$	$3x^2$
$\text{sen}(x)$	$\text{cos}(x)$
$\text{cos}(x)$	$-\text{sen}(x)$
$\text{exp}(x)$	$\text{exp}(x)$
$\ln(x)$	$1/ x $

Podemos, também, implementar o gradiente de algumas funções de várias variáveis

$f(x, y, z, \dots)$	$\nabla f(x, y, z, \dots)$
$x + y$	$(1, 1)$
$x + y + z$	$(1, 1, 1)$
$xy$	$(y, x)$

As diferenciais de funções vetoriais podem ser calculadas através dos gradientes de cada componente.

A partir dessas derivadas fundamentais, podemos calcular a derivada da função  $h(x)$  definida em (5.4). Essa função, representada na árvore [Figura 5.1](#), pode ser, também, representada, de forma paralela em relação aos ramos da árvore, como uma função composta, para facilitar a descrição matemática,

$$h(x) = f_5(f_4(f_3(f_2(f_1(x)))))$$

onde

$$\begin{aligned} f_1(x) &= (x, \text{cos}(x), x^3) \\ f_2(p, q, r) &= (p, \text{exp}(q), 1 + r), \\ f_3(u, v, w) &= (2u, 2v, \ln(w)), \\ f_4(a, b, c) &= a + b + c, \\ f_5(z) &= \text{sen}(z), \end{aligned}$$

Essa composição pode ser visualizada na [Figura 5.2](#).

Cada diferencial pode ser calculada a partir das diferenciais elementares que a linguagem já tem implementada, ilustradas nas tabelas acima.

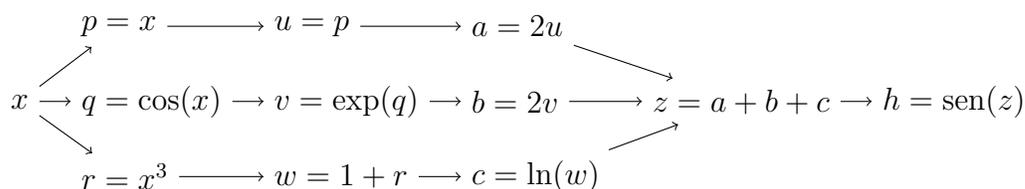


FIGURA 5.2. Ilustração da composição  $h(x) = f_5(f_4(f_3(f_2(f_1(x))))))$ .

O processo, então, é o seguinte. Uma vez fixado um ponto  $x_0$  para o cálculo de  $h'(x_0)$ , calculamos, iterativamente, os valores das funções envolvidas na composição,

$$\begin{aligned}
 (p_0, q_0, z_0) &= f_1(x_0) = (x, \cos(x), x^3) \\
 (u_0, v_0, w_0) &= f_2(p_0, q_0, r_0) = (p_0, \exp(q_0), 1 + r_0), \\
 (a_0, b_0, c_0) &= f_3(u_0, v_0, w_0) = (2u_0, 2v_0, \ln(w_0)), \\
 z_0 &= f_4(a_0, b_0, c_0) = a_0 + b_0 + c_0, \\
 h_0 &= f_5(z_0) = \text{sen}(z_0),
 \end{aligned}$$

e as diferenciais em cada ponto correspondente,

$$\begin{aligned}
 Df_1(x_0) &= \begin{bmatrix} 1 \\ -\text{sen}(x_0) \\ 3x_0^2 \end{bmatrix} \\
 Df_2(p_0, q_0, r_0) &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \exp(q_0) & 0 \\ 0 & 0 & 1 \end{bmatrix}, \\
 Df_3(u_0, v_0, w_0) &= \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & \frac{1}{|z_0|} \end{bmatrix}, \\
 Df_4(a_0, b_0, c_0) &= [1, 1, 1], \\
 Df_5(z_0) &= \cos(z_0),
 \end{aligned}$$

Finalmente, a derivada de  $h(x)$  é encontrada via composição:

$$\begin{aligned}
h'(x_0) &= Df_5(z_0)Df_4(a_0, b_0, z_0)Df_3(u_0, v_0, w_0)Df_2(p_0, q_0, z_0)Df_1(x_0) \\
&= \cos(z_0)[1, 1, 1] \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & \frac{1}{|z_0|} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \exp(q_0) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -\operatorname{sen}(x_0) \\ 3x_0^2 \end{bmatrix} \\
&= \cos(z_0)[2, 2, \frac{1}{|z_0|}] \begin{bmatrix} 1 & 0 & 0 \\ 0 & \exp(q_0) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -\operatorname{sen}(x_0) \\ 3x_0^2 \end{bmatrix} \\
&= \cos(z_0)[2, 2\exp(q_0), \frac{1}{|z_0|}] \begin{bmatrix} 1 \\ -\operatorname{sen}(x_0) \\ 3x_0^2 \end{bmatrix} \\
&= \cos(z_0) \left( 2 - 2\exp(q_0)\operatorname{sen}(x_0) + \frac{3x_0^2}{|z_0|} \right)
\end{aligned}$$

Cada cálculo desse é muito fácil e rápido para o computador. A única questão é ter todas as primitivas necessárias implementadas. E isso envolve primitivas das implementações computacionais, como `for`, `if`, `ifelse`, etc.

O cálculo pode ser acelerado quanto mais primitivas forem incluídas, para reduzir o quantidade de composições. Como boa parte da computação científica se baseia em composição de várias operações lineares, é importante implementar as diferenciais de operações envolvendo essas transformações, como cálculo do determinante, do traço, de transformação inversa, de transposta e de diversas fatorações [5].

Por exemplo, se  $T = T_{\mathbf{A}} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  é a transformação linear  $T_{\mathbf{A}}(\mathbf{u}) = \mathbf{A}\mathbf{u}$  associada a uma matriz constante  $\mathbf{A} \in \mathbb{R}^{m \times n}$ , então a sua diferencial é a própria transformação,

$$DT_{\mathbf{A}}(\mathbf{u}) = \mathbf{A} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m).$$

Se a própria matriz  $\mathbf{A} = \mathbf{A}(\mathbf{u})$  depende da variável  $\mathbf{u}$ , então precisamos calcular a diferencial  $D\mathbf{A}(\mathbf{u})$ , que vai variar com a dependência de  $\mathbf{A}(\mathbf{u})$  em  $\mathbf{u}$ . Mas se em cima disso tivermos uma função de  $\mathbf{A}$ , como  $\mathbf{A}^2$ , então precisamos calcular a diferencial dessa dependência, assim como no caso de mais uma transformação linear. Por exemplo, a função  $(\mathbf{A}, \mathbf{B}) \in \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n} \mapsto \mathbf{A}\mathbf{B} \in \mathbb{R}^{n \times n}$  é uma função bilinear nas matrizes e a sua diferencial pode ser expressa por

$$D(\mathbf{A}\mathbf{B}) = D\mathbf{A}\mathbf{B} + \mathbf{A}D\mathbf{B}.$$

Isso vem, de fato, da bilinearidade, visto que

$$(\mathbf{A} + D\mathbf{A})(\mathbf{B} + D\mathbf{B}) = \mathbf{A}\mathbf{B} + D\mathbf{A}\mathbf{B} + \mathbf{A}D\mathbf{B} + D\mathbf{A}D\mathbf{B},$$

de modo que

$$(\mathbf{A} + \mathbf{DA})(\mathbf{B} + \mathbf{DB}) - \mathbf{AB} = \mathbf{DA B} + \mathbf{A DB} + \mathcal{O}(2).$$

O termo de ordem linear em  $\mathbf{DA}$  e  $\mathbf{DB}$  (i.e. de ordem  $\mathcal{O}(1)$ ) nessa aproximação é  $\mathbf{DA B} + \mathbf{A DB}$ , que é, portanto, a diferencial do produto das matrizes.

No caso da inversa  $\mathbf{A}^{-1}$ , podemos usar que

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

para deduzir, via fórmula para a diferencial do produto matricial feita acima, que

$$\mathbf{DA}^{-1}\mathbf{A} + \mathbf{A}^{-1}\mathbf{DA} = \mathbf{0},$$

de modo que

$$\mathbf{DA}^{-1} = -\mathbf{A}^{-1}\mathbf{DA}\mathbf{A}^{-1}.$$

Para mais detalhes e para mais fórmulas, veja, por exemplo, [5].

**5.3. Análise dimensional e o Teorema de Buckingham-Pi.** A demonstração do Teorema de Buckingham-Pi é uma aplicação interessante do [Teorema 2.1](#) do Núcleo e da Imagem. Vejamos do que trata esse teorema, o seu enunciado e uma demonstração de parte dele.

A análise das dimensões físicas (ou de outra natureza) das quantidades características de um determinado problema é de grande valia em diversos aspectos. A sua aparente simplicidade revela simetrias que muitas vezes afloram de maneira observável no fenômeno em questão. Por detrás da simples análise, se escondem soluções particulares que têm papel significativo no comportamento do sistema e levam a relações aparentemente universais entre as quantidades envolvidas. Em um aspecto mais mundano, a análise dimensional é útil na verificação da corretude de certas fórmulas e de certas estimativas obtidas de um certo modelo.

No cerne do problema está um modelo que envolve certas quantidades físicas (e.g. peso, posição e velocidade de um objeto) que dependem de uma ou mais dimensões físicas (e.g. massa, comprimento e tempo). Dependendo da situação, é possível extrair quantidades adimensionais que determinam relações importantes entre essas quantidades.

Para ilustrar o processo, vamos considerar o modelo clássico de um pêndulo. Nesse modelo, temos uma fina haste de comprimento  $\ell$  e massa considerada desprezível, com uma das extremidades fixas e com a outra extremidade prendendo um objeto de massa  $m$ . Isso sob a ação da força gravitacional, com aceleração gravitacional  $g$ . Um pequeno movimento na massa faz com que o pêndulo oscile, com um período  $\tau$ , descrevendo um arco de círculo a cada oscilação, com ângulo de oscilação igual a  $2\theta$ ,

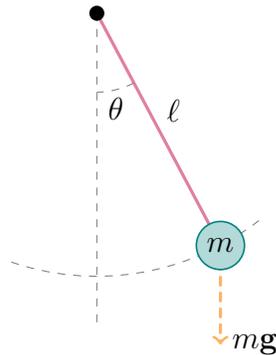


FIGURA 5.3. Ilustração de um pêndulo com haste de comprimento  $\ell$ , objeto com massa  $m$ , sob a ação da gravidade com aceleração  $g$ , oscilando até um ângulo máximo  $\theta$  em relação ao eixo vertical e com o período de oscilação  $\tau$ , não representado.

ou seja,  $\theta$  é o ângulo entre o eixo vertical direcionado para baixo e o vetor posição do ponto de maior distância do eixo, conforme ilustrado na [Figura 5.3](#).

Temos, com isso, cinco quantidades físicas. As dimensões físicas envolvidas são a massa  $M$ , o comprimento  $L$  e o tempo  $T$ . Denotando  $[q]$  a dimensão física de uma quantidade qualquer  $q$ , temos

$$\begin{aligned}[\ell] &= L, \\ [m] &= M, \\ [g] &= \frac{L}{T^2}, \\ [\tau] &= T, \\ [\theta] &= 1.\end{aligned}$$

Observe que  $\theta$  é adimensional, o que é denotado por dimensão 1, pois é definido por uma razão  $\theta = a/r$  entre um arco de comprimento  $a$  de uma circunferência e o raio  $r$  da mesma, ambas com a mesma dimensão, de forma que  $\theta$  dimensão tem dimensão

$$[\theta] = \left[\frac{a}{r}\right] = \frac{[a]}{[r]} = \frac{L}{L} = 1.$$

Olhando para uma combinação qualquer de potências dessas quantidades  $\ell$ ,  $m$ ,  $g$ ,  $\tau$  e  $\theta$ , podemos considerar uma nova quantidade que chamamos de  $\Pi$ , símbolo usado por Buckingham e que deu nome ao resultado hoje conhecido como *Teorema*

de *Buckingham-Pi*. Mais precisamente, considerar uma nova quantidade  $\Pi$  dada por

$$\Pi = \ell^a m^b g^c \tau^d \theta^e, \quad (5.5)$$

onde  $a, b, c, d, e$  são, em geral, números racionais, mas, em teoria, podem ser quaisquer números reais, desde que a potência em questão faça sentido.

Olhando para a dimensão de  $\Pi$ , obtemos

$$[\Pi] = [\ell^a m^b g^c \tau^d \theta^e] = [\ell^a][m^b][g^c][\tau^d][\theta^e] = \mathbf{L}^a \mathbf{M}^b \left(\frac{\mathbf{L}}{\mathbf{T}^2}\right)^c \mathbf{T}^d \mathbf{1}^e.$$

Combinando os termos com a mesma dimensão, obtemos

$$[\Pi] = \mathbf{L}^{a+c} \mathbf{M}^b \mathbf{T}^{d-2c}. \quad (5.6)$$

A partir disso, buscamos combinações que nos dão uma quantidade  $\Pi$  que seja *adimensional*. Para isso, cada potência acima deve ser igual a zero. Ou seja, devemos resolver o sistema linear

$$\begin{cases} a + c = 0 \\ b = 0 \\ d - 2c = 0 \\ e \text{ qualquer} \end{cases} \quad (5.7)$$

Esse é um sistema homogêneo, de três equações com cinco incógnitas. Se não houver nenhuma redundância, esperamos um conjunto solução de dimensão dois. De fato, resolvendo, obtemos o conjunto solução

$$\begin{aligned} C &= \{(a, b, c, d, e) = (r, 0, -r, -2r, s); r, s \in \mathbb{R}\} \\ &= \text{span}\{(1, 0, -1, -2, 0), (0, 0, 0, 0, 1)\}. \end{aligned}$$

Usando cada vetor dessa base, encontramos duas quantidades adimensionais, denotadas por  $\Pi_1$  e  $\Pi_2$ :

$$\Pi_1 = \ell^1 m^0 g^{-1} \tau^{-2} \theta^0 = \frac{\ell}{g\tau^2}, \quad \Pi_2 = \ell^0 m^0 g^0 \tau^0 \theta^1 = \theta. \quad (5.8)$$

Observe que a segunda é o próprio ângulo  $\theta$ , que já sabíamos ser adimensional desde o início. Mas agora conhecemos também uma outra quantidade adimensional,  $\Pi_1$ . Mas a escolha da base é arbitrária. Por exemplo, qualquer potência  $\Pi_i^r$  de  $\Pi_i$ ,  $i = 1, 2$ , continua sendo adimensional, assim como qualquer combinação  $\Pi_1^{r_1} \Pi_2^{r_2}$ .

A partir dessa quantidade  $\Pi_1$ , obtemos uma relação entre o período  $\tau$  e as outras quantidades:

$$\tau \propto \sqrt{\frac{\ell}{g}}.$$

Uma análise mais formal, nos leva a uma relação  $f(\Pi_1, \Pi_2) = 0$  entre as quantidades adimensionais que se refletem em uma dependência do período no ângulo  $\theta$ , da forma

$$\theta = g(\theta) \sqrt{\frac{\ell}{g}}.$$

Para pequenas oscilações, isso é bem aproximado por uma constante de proporcionalidade. Por outro ponto de vista, para ângulos pequenos de oscilação, podemos aproximar o problema do pêndulo por um de massa mola, para o qual temos uma solução simples explícita, que nos dá a relação

$$\theta = 2\pi \sqrt{\frac{g}{\ell}}.$$

Mas para ângulos maiores, a dependência em  $\theta$  não pode ser desprezada.

No exemplo acima, reduzimos o problema de encontrar quantidades adimensionais a um sistema linear homogêneo, que resolvemos explicitamente. Em geral, o que podemos usar é o **Teorema 2.1** do Núcleo e da Imagem para, pelo menos, responder quantas quantidades adimensionais podemos esperar extrair do modelo. Isso pode ser formulado de seguinte maneira.

**TEOREMA 5.1.** *Considere um sistema com  $n$  quantidades  $q_1, \dots, q_n$  em que  $m$  dimensões fundamentais estão envolvidas. Se  $n > m$ , então pelo menos  $n - m$  grandezas adimensionais  $\Pi_1, \dots, \Pi_{n-m}$  independentes podem ser definidas como potências das quantidades originais. Além disso, cada equação (escalar)  $f(q_1, \dots, q_n) = 0$  entre essas quantidades, pode ser substituída por uma relação correspondente  $f^*(\Pi_1, \dots, \Pi_{n-m}) = 0$  entre os grupos adimensionais  $\Pi_i$*

Esse teorema foi proposto em 1914 por Buckingham. O nome Pi se deve à notação matemática  $\Pi$  para um produtório, conforme originalmente denotado por Buckingham, visto que as quantidades adimensionais são produtos das quantidades originais do modelo.

Os fundamentos desse resultado de análise dimensional repousam, essencialmente, em duas hipóteses:

- (1) *Universalidade*: todas as variáveis relevantes foram incluídas no modelo.
- (2) *Homogeneidade dimensional*: A relação física  $f(q_1, \dots, q_n) = 0$  proposta pelo modelo é dimensionalmente homogênea.

No enunciado do teorema, observe que afirmamos a existência de, pelo menos,  $n - m$  quantidades adimensionais *independentes*. Isso porque, dada uma quantidade adimensional  $\Pi$ , qualquer potência  $\Pi^r$  também é adimensional. Mais geralmente,

dadas quaisquer quantidades adimensionais  $\Pi_1, \dots, \Pi_k$ , uma combinação qualquer da forma  $\Pi_1^{r_1} \dots \Pi_k^{r_k}$  também é adimensional. Portanto, quando dizemos que temos  $n - m$  quantidades adimensionais  $\Pi_1, \dots, \Pi_k$ , queremos dizer que nenhuma delas é combinação multiplicativa de potências das outras quantidades. Observe que tomando o logaritmo, podemos reinterpretar isso em termos de combinações lineares. Definir o logaritmo de quantidades dimensionais, satisfazendo as propriedades esperadas de um logaritmo como o de números reais, é possível mas requer um certo formalismo, que não faremos aqui.

Além disso, não vamos, aqui, tornar preciso o que são “dimensões” e “quantidades”, no contexto do teorema, nem o que são essas condições de universalidade e homogeneidade, mas saibam que tudo isso pode ser devidamente formalizado com estruturas matemáticas. Para mais detalhes, vejam, por exemplo [2, 6, 14].

Mas podemos dar detalhes da demonstração, pelo menos da primeira parte, envolvendo apenas a existência das quantidades adimensionais. A obtenção da relação adimensional  $f^*(\Pi_1, \dots, \Pi_{n-m}) = 0$  é um pouco mais delicada.

DEMONSTRAÇÃO. Formemos o produto

$$\Pi = q_1^{r_1} \dots q_n^{r_n}.$$

Queremos saber para quais escolhas de  $r_i$  o produto é adimensional. Para ver isso, olhamos para a dimensão do produto e substituímos cada  $[q_i]$  em termos de suas dimensões fundamentais. Se temos apenas  $m$  dimensões fundamentais  $D_1, \dots, D_m$ , então essa substituição dá origem a um produto

$$[\Pi] = [q_1]^{r_1} \dots [q_n]^{r_n} = D_1^{s_1} \dots D_m^{s_m}, \quad (5.9)$$

onde cada  $s_j$ ,  $j = 1, \dots, m$ , é uma combinação linear dos  $r_i$ ,  $i = 1, \dots, n$ .

Mais precisamente, escrevamos a dimensão de cada  $q_i$  na forma

$$[q_i] = D_1^{a_{i,1}} D_2^{a_{i,2}} \dots D_m^{a_{i,m}}.$$

Multiplicando as potências de  $q_i$ , obtemos a dimensão

$$[\Pi] = (D_1^{a_{1,1}} D_2^{a_{1,2}} \dots D_m^{a_{1,m}})^{r_1} (D_1^{a_{2,1}} D_2^{a_{2,2}} \dots D_m^{a_{2,m}})^{r_2} \dots (D_1^{a_{n,1}} D_2^{a_{n,2}} \dots D_m^{a_{n,m}})^{r_n}.$$

Reescrevendo em termos de potência de cada dimensão, obtemos

$$[\Pi] = D_1^{a_{1,1}r_1 + \dots + a_{n,1}r_n} D_2^{a_{1,2}r_1 + \dots + a_{n,2}r_n} \dots D_m^{a_{1,m}r_1 + \dots + a_{n,m}r_n}.$$

Isso pode ser escrito na forma

$$[\Pi] = D_1^{s_1} D_2^{s_2} \dots D_m^{s_m},$$

onde

$$\begin{cases} s_1 = a_{1,1}r_1 + \cdots + a_{n,1}r_n, \\ s_2 = a_{1,2}r_1 + \cdots + a_{n,2}r_n, \\ \cdots \\ s_n = a_{1,m}r_1 + \cdots + a_{n,m}r_n. \end{cases}$$

Isso é exatamente (5.9), com as combinações lineares de  $s_i$  em termos dos  $r_j$  tornada mais explícita.

Escrevemos essa dependência linear dos  $s_i$  nos  $r_j$  como

$$\mathbf{s} = \mathbf{A}\mathbf{r},$$

onde  $\mathbf{s} = (s_1, \dots, s_m) \in \mathbb{R}^m$ ,  $\mathbf{r} = (r_1, \dots, r_n) \in \mathbb{R}^n$  e  $\mathbf{A} = (a_{ij})_{ij} \in \mathbb{R}^{m \times n}$  é uma matriz  $m \times n$ .

Para que o produto  $[\Pi] = D_1^{s_1} \cdots D_m^{s_m}$  seja adimensional, devemos ter cada potência se anulando, i.e.

$$s_i = 0, \quad i = 1, \dots, m.$$

Isso nos leva a um sistema linear homogêneo de  $m$  equações lineares para  $n$  incógnitas  $r_1, \dots, r_n$ ,

$$\mathbf{A}\mathbf{r} = \mathbf{0}.$$

Ou seja, reduzimos o problema a se determinar o núcleo da transformação linear associada a uma matriz de tamanho  $m \times n$ , ou seja

$$\ker(\mathbf{A}) = \{\mathbf{r} \in \mathbb{R}^n; \mathbf{A}\mathbf{r} = \mathbf{0}\}.$$

Pelo Teorema 2.1 do Núcleo e da Imagem,

$$\dim \ker(\mathbf{A}) + \dim \operatorname{Im}(\mathbf{A}) = n.$$

Como  $\dim \operatorname{Im}(\mathbf{A}) \leq m$ , obtemos

$$\dim \ker(\mathbf{A}) = n - \dim \operatorname{Im}(\mathbf{A}) \geq n - m.$$

Logo, a dimensão do núcleo é no mínimo  $n - m$ . Portanto, temos, no mínimo,  $n - m$  soluções linearmente independentes,  $\mathbf{r}^k$ ,  $k = 1, \dots, n - m$ .

Cada solução  $\mathbf{r}^k = (r_1^k, \dots, r_n^k)$ ,  $k = 1, \dots, n - m$ , nos dá uma quantidade adimensional

$$\Pi_k = q_1^{r_1^k} \cdots q_n^{r_n^k}.$$

Isso corresponde, justamente, às  $n - m$  grandezas adimensionais do enunciado do teorema. Pelo menos. A matriz pode não ter posto máximo, de forma que o núcleo pode ser maior do que  $n - m$ , revelando outras quantidades adimensionais.  $\square$

**5.4. Dualidade em programação linear.** Considere um problema de otimização linear da forma

$$\max_{\substack{\mathbf{u} \geq \mathbf{0} \\ \mathbf{A}\mathbf{u} \leq \mathbf{b}}} \langle \mathbf{f}, \mathbf{u} \rangle, \quad (5.10)$$

onde  $\mathbf{A} : X \rightarrow Y$  é uma transformação linear entre espaços vetoriais  $X$  e  $Y$  sobre um mesmo corpo  $\mathbb{K}$ ;  $\mathbf{b} \in Y$ ; e  $\mathbf{f} \in X^*$  são fornecidos; e  $\geq$  são relações de ordem parcial em  $X$  e  $Y$  (i.e.  $\mathbf{u} \geq \mathbf{0}$  em  $X$  e  $\mathbf{A}\mathbf{u} \leq \mathbf{b}$  em  $Y$ ), compatíveis com as estruturas vetoriais correspondentes (veja [Exercício 5.13](#)). A maximização ocorre em relação à variável  $\mathbf{u} \in X$ , restrita ao conjunto admissível

$$\mathcal{U} = \{\mathbf{u} \geq \mathbf{0}; \mathbf{A}\mathbf{u} \leq \mathbf{b}\}. \quad (5.11)$$

A formulação (5.10), por ser a original, pode ser chamada de *problema primal*. O *problema dual* tem a forma

$$\min_{\substack{\mathbf{g} \geq \mathbf{0} \\ \mathbf{A}^*\mathbf{g} \geq \mathbf{f}}} \langle \mathbf{g}, \mathbf{b} \rangle, \quad (5.12)$$

onde a minimização, agora, é na variável  $\mathbf{g} \in X^*$ , restrita ao conjunto admissível

$$\mathcal{F} = \{\mathbf{g} \in Y^*; \mathbf{A}^*\mathbf{g} \geq \mathbf{f}\}, \quad (5.13)$$

e onde as relações são as relações de ordem parcial induzidas nos duais  $X^*$  e  $Y^*$  (veja [Exercício 7.11](#)).

Qual a relação entre os dois problemas? Ela é dada pelo seguinte resultado, conhecido como Teorema Fraco de Dualidade.

**TEOREMA 5.2 (Fraco de Dualidade).** *Sejam  $X$  e  $Y$  espaços vetoriais sobre um mesmo corpo com relação de ordem parcial  $\leq$  (mesma notação para os dois espaços),  $\mathbf{A} : X \rightarrow Y$  uma transformação linear,  $\mathbf{b} \in Y$  e  $\mathbf{f} \in X^*$  são fornecidos. Supondo que os conjuntos admissíveis  $\mathcal{U}$  e  $\mathcal{F}$ , definidos em (5.11) e (5.13) sejam não vazios, vale a desigualdade*

$$\sup_{\substack{\mathbf{u} \geq \mathbf{0} \\ \mathbf{A}\mathbf{u} \leq \mathbf{b}}} \langle \mathbf{f}, \mathbf{u} \rangle \leq \inf_{\substack{\mathbf{g} \geq \mathbf{0} \\ \mathbf{A}^*\mathbf{g} \geq \mathbf{f}}} \langle \mathbf{g}, \mathbf{b} \rangle. \quad (5.14)$$

**DEMONSTRAÇÃO.** Observe que para todo  $\mathbf{u} \in \mathcal{U}$  e para todo  $\mathbf{g} \in \mathcal{F}$ , vale

$$\langle \mathbf{f}, \mathbf{u} \rangle \leq \langle \mathbf{A}^*\mathbf{g}, \mathbf{u} \rangle = \langle \mathbf{g}, \mathbf{A}\mathbf{u} \rangle \leq \langle \mathbf{g}, \mathbf{b} \rangle,$$

de modo que

$$\sup_{\substack{\mathbf{u} \geq \mathbf{0} \\ \mathbf{A}\mathbf{u} \leq \mathbf{b}}} \langle \mathbf{f}, \mathbf{u} \rangle \leq \inf_{\substack{\mathbf{g} \geq \mathbf{0} \\ \mathbf{A}^*\mathbf{g} \geq \mathbf{f}}} \langle \mathbf{g}, \mathbf{b} \rangle.$$

□

COROLÁRIO 5.1. *Se*

$$\inf_{\substack{\mathbf{g} \geq \mathbf{0} \\ \mathbf{A}^* \mathbf{g} \geq \mathbf{f}}} \langle \mathbf{g}, \mathbf{b} \rangle = -\infty,$$

*então o problema primal é inviável (não possui solução). Por outro lado, se*

$$\sup_{\substack{\mathbf{u} \geq \mathbf{0} \\ \mathbf{A}\mathbf{u} \leq \mathbf{b}}} \langle \mathbf{f}, \mathbf{u} \rangle = \infty,$$

*então o problema dual é inviável.*

COROLÁRIO 5.2. *Se  $\mathbf{u} \in \mathcal{U}$  e  $\mathbf{g} \in \mathcal{F}$  admissíveis são tais que*

$$\langle \mathbf{g}, \mathbf{b} \rangle = \langle \mathbf{f}, \mathbf{u} \rangle,$$

*então  $\mathbf{u}$  e  $\mathbf{g}$  são soluções ótimas dos problemas primal (5.10) e dual (5.12), respectivamente.*

Sob certas condições, o máximo e o mínimo acima coincidem, que é a versão forte do teorema de dualidade, ou Teorema Forte de Dualidade.

EXEMPLO 5.1. Não há nada de tão especial na forma primal. Na verdade, dependendo do contexto, se o interesse inicial for pelo problema (5.12), então podemos chamá-lo de problema primal, de forma que (5.10) seria, então, o seu dual.

EXEMPLO 5.2. Também podemos considerar inicialmente uma variação do (5.10), na forma, muito comum,

$$\min_{\substack{\mathbf{u} \geq \mathbf{0} \\ \mathbf{A}\mathbf{u} \leq \mathbf{b}}} \langle \mathbf{f}, \mathbf{u} \rangle. \quad (5.15)$$

Esse problema pode ser reescrito como

$$\max_{\substack{\mathbf{u} \geq \mathbf{0} \\ \mathbf{A}\mathbf{u} \leq \mathbf{b}}} \langle -\mathbf{f}, \mathbf{u} \rangle, \quad (5.16)$$

que é, sim, da forma (5.10).

### 5.5. Problemas primal e dual em máquina de vetores de suporte.

## 6. Exercícios

### Exercícios

- 6.1.** Sejam  $X$  e  $Y$  espaços vetoriais sobre um mesmo corpo e seja  $J : X \rightarrow Y$  um isomorfismo entre  $X$  e  $Y$ , i.e. uma transformação linear que é uma bijeção entre o seu domínio e o seu contra-domínio. Por ser uma bijeção, a sua inversa  $J^{-1} : Y \rightarrow X$  está bem definida e satisfaz  $J^{-1}(J(\mathbf{u})) = \mathbf{u}$ , para todo  $\mathbf{u} \in X$  e  $J(J^{-1}(\mathbf{v})) = \mathbf{v}$ , para todo  $\mathbf{v} \in Y$ . Mostre que  $J^{-1} : Y \rightarrow X$  também é linear.
- 6.2.** Sejam  $T : X \rightarrow Y$  e  $S : Y \rightarrow Z$  transformações lineares invertíveis, onde  $X$ ,  $Y$  e  $Z$  são espaços vetoriais sobre o mesmo corpo. Mostre que  $(S \circ T) : X \rightarrow Z$  também é invertível e a sua inversa é dada por  $(S \circ T)^{-1} = T^{-1} \circ S^{-1}$ .
- 6.3.** Seja  $T : X \rightarrow Y$  uma transformação linear entre espaços vetoriais  $X$  e  $Y$  sobre o mesmo corpo, com  $X$  de dimensão finita. Mostre que  $X/\ker(T)$  é isomorfo a  $\text{Im}(T)$  (veja [Teorema 2.1](#)).
- 6.4.** Sejam  $X, Y$  e  $Z$  espaços vetoriais sobre um mesmo corpo,  $T, L \in \mathcal{L}(X, Y)$  e  $R, S \in \mathcal{L}(Y, Z)$  e  $\lambda \in \mathbb{K}$ . Mostre que

$$S \circ (T + \lambda L) = (S \circ T) + \lambda(T \circ L)$$

e

$$(S + \lambda R) \circ T = (S \circ T) + \lambda(R \circ T).$$

- 6.5.** O [Teorema 2.1](#) do Núcleo e da Imagem nos fornece uma outra demonstração de que se  $U, V$  são dois subespaços de um espaço vetorial  $X$ , então  $\dim(U \cup V) = \dim(U) + \dim(V) - \dim(U \cap V)$ . Para isso, considere a transformação linear  $T : U \times V \rightarrow X$  dada por  $T(\mathbf{u}, \mathbf{v}) = \mathbf{u} + \mathbf{v}$ , para  $(\mathbf{u}, \mathbf{v}) \in U \times V$ . Observe que, nessa transformação, o domínio é o produto cartesiano  $U \times V$  enquanto que  $X$  é o contra-domínio. Identifique a dimensão do domínio  $U \times V$ , assim como o núcleo de  $T$  em  $U \times V$  e a imagem de  $T$  no contradomínio  $X$  e use o Teorema do Núcleo e da Imagem para deduzir o resultado.
- 6.6.** Encontre uma base para o núcleo das seguintes transformações lineares
- (1)  $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  dado por  $T(x, y, z) = (x - y, y - z, z - x)$ ;
  - (2)  $T : \mathbb{R}^3 \rightarrow \mathbb{R}^2$  dado por  $T(x, y, z) = (x - y, y - z)$ ;
  - (3)  $T : \mathcal{P}_2 \rightarrow \mathcal{P}_2$  dado por  $T(p)(x) = p''(x)$ ;
  - (4)  $T : \mathbb{R}^{2 \times 2} \rightarrow \mathbb{R}^2$  dado por  $T\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = (a + b, d - c)$
- 6.7.** Encontre uma base para a imagem das seguintes transformações lineares
- (1)  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^3$  dado por  $T(x, y) = (x - y, x - 2y, 3x - 2y)$ ;
  - (2)  $T : \mathbb{R}^3 \rightarrow \mathbb{R}^4$  dado por  $T(x, y, z) = (x - y, y - z, x - z, x + y + z)$ ;
  - (3)  $T : \mathcal{P}_2 \rightarrow \mathcal{P}_2$  dado por  $T(p)(x) = p''(x)$ ;
  - (4)  $T : \mathbb{R}^{2 \times 2} \rightarrow \mathbb{R}^2$  dado por  $T\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = (a + b, d - c)$

- 6.8.** Encontre uma transformação linear  $T : \mathbb{R}^3 \rightarrow \mathbb{R}^2$  tal que o núcleo seja o plano  $x + y + z = 0$  no domínio  $\mathbb{R}^3$  e a imagem seja a reta  $y = x$  no contra-domínio  $\mathbb{R}^2$ .
- 6.9.** Sejam  $T : X \rightarrow Y$  e  $S : Y \rightarrow Z$  transformações lineares entre espaços vetoriais  $X, Y, Z$  sobre o mesmo corpo. Então
- (1) Mostre que  $\text{Im}(ST) \subset \text{Im}(S)$ .
  - (2) Mostre que  $\text{ker}(T) \subset \text{ker}(ST)$ .
  - (3) Mostre que  $\dim(\text{ker}(ST)) \leq \dim(\text{ker}(S)) + \dim(\text{ker}(T))$ .
  - (4) Dê um exemplo em que vale a igualdade  $\dim(\text{ker}(ST)) = \dim(\text{ker}(S)) + \dim(\text{ker}(T))$ .
- 6.10.** Seja  $T : X \rightarrow Y$  uma transformação linear entre dois espaços vetoriais de dimensão finita  $X$  e  $Y$ . Mostre que existe um espaço vetorial  $Z$  e duas transformações lineares  $S : X \rightarrow Z$  e  $R : Z \rightarrow Y$  tais que  $T = RS$ , com  $S$  sobrejetiva e  $R$  injetiva.
- 6.11.** Sejam  $T : X \rightarrow Y$  e  $S : Y \rightarrow Z$  transformações lineares entre espaços vetoriais  $X, Y, Z$  sobre um mesmo corpo.
- (1) Suponha que  $T$  e  $S$  sejam sobrejetivos. Mostre que  $ST$  é sobrejetivo.
  - (2) Suponha que  $T$  e  $S$  sejam injetivos. Mostre que  $ST$  é injetivo.
  - (3) Suponha que  $T$  e  $S$  sejam bijetivos. Mostre que  $ST$  é bijetivo.
- 6.12.** Sejam  $T : X \rightarrow Y$  e  $S : Y \rightarrow Z$  transformações lineares entre espaços vetoriais  $X, Y, Z$  sobre um mesmo corpo. Mostre que se  $ST : X \rightarrow Z$  é bijetiva, então  $T$  é injetiva e  $S$  é sobrejetiva. Dê um exemplo mostrando que não é necessário que  $T$  seja sobrejetiva e que  $S$  seja injetiva, para que  $ST$  seja bijetiva.
- 6.13.** Seja  $T : X \rightarrow Y$  uma transformação linear entre dois espaços vetoriais  $X$  e  $Y$  de dimensão finita.
- (1) Mostre que  $T$  é sobrejetiva se, e somente se, leva um conjunto de vetores  $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  que geram  $X$  em um conjunto  $\{T\mathbf{w}_1, \dots, T\mathbf{w}_n\}$  de vetores que geram  $Y$ .
  - (2) Mostre que  $T$  é injetiva se, e somente se, leva um conjunto de vetores  $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  LI em  $X$  em um conjunto de vetores  $\{T\mathbf{w}_1, \dots, T\mathbf{w}_n\}$  LI em  $Y$ .
  - (3) Mostre que  $T$  é bijetiva se, e somente se, leva uma base  $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  de  $X$  em uma base  $\{T\mathbf{w}_1, \dots, T\mathbf{w}_n\}$  de  $Y$ .
- 6.14.** Seja  $T : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$  a transformação linear que leva uma matriz quadrada  $\mathbf{A} \in \mathbb{R}^{n \times n}$  em sua simetrização  $T(\mathbf{A}) = (\mathbf{A} + \mathbf{A}^{\text{tr}})/2$ , onde  $\mathbf{A}^{\text{tr}}$  é a sua transposta (veja [Exemplo 1.23](#)). Encontre o posto  $\dim(\text{Im}(T))$  de  $T$  e caracterize o seu núcleo  $\text{ker}(T)$ .

- 6.15.** Seja  $T : X \rightarrow Y$  uma transformação linear entre dois espaços vetoriais  $X$  e  $Y$  sobre o mesmo corpo  $\mathbb{K}$ . Seja  $C \subset X$  um conjunto convexo no domínio de  $T$ . Mostre que a imagem  $TC = \{T\mathbf{u}; \mathbf{u} \in C\}$  de  $C$  sob  $T$  é um conjunto convexo no contra-domínio  $Y$ .
- 6.16.** Seja  $T : X \rightarrow Y$  uma transformação linear entre dois espaços vetoriais  $X$  e  $Y$  sobre o mesmo corpo  $\mathbb{K}$ . Seja  $C \subset Y$  um conjunto convexo no contra-domínio. Mostre que a pré-imagem  $T^{-1}C = \{\mathbf{u} \in X; T\mathbf{u} \in C\}$  é um conjunto convexo no domínio  $X$ .
- 6.17.** Seja  $M : X \rightarrow Y$  uma transformação invertível, entre dois espaços vetoriais  $X$  e  $Y$  de mesma dimensão finita. Considere uma base qualquer  $\mathcal{b}$  de  $X$ . Encontre uma base  $\mathcal{a}$  tal que  $[M]_{\mathcal{b}}^{\mathcal{a}} = \mathbf{I}$ , onde  $\mathbf{I}$  é a matriz identidade.
- 6.18.** Considere uma corda de comprimento  $\ell$  com uma extremidade fixa e a outra presa a um objeto de massa  $m$ . Suponha que esse sistema corda-massa gire em um movimento circular uniforme, com velocidade tangencial constante  $v$ . Esse movimento gera uma tensão  $T$  na corda. A tensão é uma força. Use análise dimensional para deduzir a relação

$$\frac{T\ell}{mv^2} = \text{constante.}$$

- 6.19.** Em 1941, G. I. Taylor, um famoso físico e matemático britânico, com grandes contribuições à mecânica dos fluidos, fez uma [estimativa do rendimento da primeira bomba nuclear](#) detonada pelos EUA, na chamada experiência Trinity. A estimativa foi feita apenas com análise dimensional e [dados da evolução da frente de onda obtidos de uma filmagem da explosão](#). Assumindo que os únicos parâmetros relevantes para o cálculo do rendimento  $E$  (energia despendida) são a densidade  $\rho$  do ar, o raio da onda de choque  $r = r(t)$  e o instante de tempo  $t$ , encontre uma relação adimensional entre essas quantidades.
- 6.20.** Estimar o atrito na parede de um tubo é uma das tarefas mais comuns em aplicações de engenharia envolvendo fluidos. Para tubos longos circulares e ásperos em regime de escoamento turbulento, a tensão de cisalhamento na parede,  $\tau_w$ , é uma função da densidade  $\rho$ , viscosidade (dinâmica)  $\mu$ , velocidade média  $V$ , diâmetro do cano  $d$ , e da altura de rugosidade da parede  $\varepsilon$ . Assim, funcionalmente podemos escrever a relação  $\tau_w = f(\rho, \mu, V, d, \varepsilon)$ .
- (1) Usando o procedimento do Teorema de Buckingham-Pi, mostre como chegar nos grandezas adimensionais

$$C_f = \frac{\tau_w}{\rho V^2}, \quad \text{Re} = \frac{\rho V d}{\mu}, \quad \varepsilon^* = \frac{\varepsilon}{d},$$

(2) Conclua que a relação pode ser simplificada para  $C_f = f^*(Re, \varepsilon^*)$ .  
O termo  $C_f$  é chamado de *coeficiente de atrito na parede*;  $Re$  é chamado de *número de Reynolds*; e  $\varepsilon^*$  é chamado de *rugosidade relativa*.



## CAPÍTULO 4

### Operadores Lineares

Quando uma transformação linear tem o seu contra-domínio igual ao seu domínio, dizemos que a transformação é um *operador* linear. Operadores lineares trazem uma riqueza de propriedades e aplicações, que começaremos a explorar, aqui.

#### 1. Fundamentos

##### 1.1. Definição de operador linear.

DEFINIÇÃO 1.1. *Dado um espaço vetorial  $X$ , uma transformação linear  $T : X \rightarrow X$  desse espaço nele mesmo leva o nome especial de **operador linear** em  $X$ .*

**1.2. A álgebra dos operadores lineares.** No caso em que o domínio e o contra-domínio são o mesmo espaço vetorial, podemos compor transformações lineares em qualquer ordem e repetidamente. Além disso, temos uma estrutura de *álgebra associativa*.

PROPOSIÇÃO 1.1. *Dado um espaço vetorial  $X$  sobre um corpo  $\mathbb{K}$ , o espaço  $\mathcal{L}(X, X)$  dos operadores lineares em  $X$  é denotado simplesmente por  $\mathcal{L}(X)$  e forma uma álgebra em relação às estruturas de espaço vetorial em  $\mathcal{L}(X)$  e de composição de operadores, ou seja,  $\mathcal{L}(X)$  é um espaço vetorial sobre  $\mathbb{K}$  e a composição define uma operação binária  $\circ : \mathcal{L}(X) \times \mathcal{L}(X) \rightarrow \mathcal{L}(X)$  que é distributiva à direita e à esquerda e é compatível com o corpo, no sentido de que  $(\lambda S) \circ (\mu T) = \lambda \mu (S \circ T)$ . Mais ainda,  $\mathcal{L}(X)$  é uma álgebra associativa, visto que a composição é associativa (mas não é comutativa).*

DEMONSTRAÇÃO. Segue da **Proposição 3.2** e da **Proposição 3.3**. □

Composições do mesmo operador ganham uma notação adequada.

DEFINIÇÃO 1.2. *Dado um espaço vetorial  $X$  e um operador linear  $T \in \mathcal{L}(X)$ , a composição de  $T$  consigo mesmo, repetidas vezes, é denotada por  $T^k = TT \cdots T$  ( $k$  vezes),  $k \in \mathbb{N}$ .*

DEFINIÇÃO 1.3. Dado um espaço vetorial  $X$  sobre um corpo  $\mathbb{K}$ , um operador linear  $\mathbb{T} \in \mathcal{L}(X)$  e um polinômio  $p = p(x) = a_0 + a_1x + \dots + a_nx^n$  com coeficientes  $a_0, \dots, a_n$  no corpo  $\mathbb{K}$ , podemos definir o operador  $p(\mathbb{T}) \in \mathcal{L}(X)$  por

$$p(\mathbb{T})\mathbf{u} = (a_0\mathbb{I} + a_1\mathbb{T} + \dots + a_n\mathbb{T}^n)\mathbf{u} = a_0\mathbf{u} + a_1\mathbb{T}\mathbf{u} + \dots + a_n\mathbb{T}^n\mathbf{u},$$

para todo  $\mathbf{u} \in X$ .

PROPOSIÇÃO 1.2. Seja  $\mathbb{T} \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  sobre um corpo  $\mathbb{K}$ . Se  $p$  e  $q$  são polinômios sobre o corpo  $\mathbb{K}$ , então vale a identidade

$$(pq)(\mathbb{T}) = p(\mathbb{T})q(\mathbb{T}).$$

Em particular, se  $p$  está na forma fatorada  $p(x) = a(x - x_0)^{n_0} \dots (x - x_m)^{n_m}$ , com  $x_0, \dots, x_m \in \mathbb{K}$ ,  $n_1, \dots, n_m \in \mathbb{N}$  e  $m \in \mathbb{N}$ , então

$$p(\mathbb{T}) = a(\mathbb{T} - x_0\mathbb{I})^{n_0} \dots (\mathbb{T} - x_m\mathbb{I})^{n_m}.$$

DEMONSTRAÇÃO. Basta expandir e fazer as devidas manipulações algébricas. Deixamos para o leitor verificar.  $\square$

OBSERVAÇÃO 1.1. Nesse caso em que  $\mathbb{T}, \mathbb{S} \in \mathcal{L}(X, X)$ , em um mesmo espaço vetorial  $X$ , podemos considerar tanto a composição  $\mathbb{S} \circ \mathbb{T}$  como a composição  $\mathbb{T} \circ \mathbb{S}$ . É natural, nesse momento, questionar se vale a comutatividade  $\mathbb{T} \circ \mathbb{S} = \mathbb{S} \circ \mathbb{T}$ . Isso não é verdade, em geral. Por exemplo, considerando  $X = \mathbb{R}^2$ ,  $\mathbb{T}(x, y) = (-y, x)$  a rotação de  $90^\circ$  no sentido trigonométrico e  $\mathbb{S}(x, y) = (x, -y)$  a reflexão no eixo  $x$ , temos que  $(\mathbb{S} \circ \mathbb{T})(x, y) = (-y, -x)$  é a reflexão em relação à reta  $y = -x$ , enquanto que  $(\mathbb{T} \circ \mathbb{S})(x, y) = (y, x)$  é a reflexão em relação à reta  $y = x$ .

OBSERVAÇÃO 1.2. Outro exemplo interessante de não-comutatividade aparece em  $X = \mathcal{C}^\infty(\mathbb{R})$  com as operações fundamentais do Cálculo. Mais precisamente, seja  $\mathbb{T}$  o operador de derivação, i.e. dado  $f \in \mathcal{C}^\infty(\mathbb{R})$ , temos  $\mathbb{T}(f)$  como sendo a derivada de  $f$ ,

$$\mathbb{T}(f)(x) = \frac{df}{dx}(x).$$

E seja  $\mathbb{S}$  o operador que leva uma função  $f \in \mathcal{C}^\infty(\mathbb{R})$  na sua primitiva que se anula na origem, i.e.

$$\mathbb{S}(f)(x) = \int_0^x f(\xi) d\xi.$$

Naturalmente,

$$(\mathbb{T} \circ \mathbb{S})(f) = f,$$

para todo  $f \in \mathcal{C}^\infty(\mathbb{R})$ . Por outro lado,

$$(\mathbb{S} \circ \mathbb{T})(f) = f - f(0),$$

para  $f \in \mathcal{C}^\infty(\mathbb{R})$  qualquer. Podemos escrever que

$$TS = I, \quad ST = I - \delta_0,$$

onde  $I$  é o operador identidade e  $\delta_0(f) = f(0)$  é a delta de Dirac.

EXEMPLO 1.1. O espaço  $\mathbb{R}^3$  munido do produto vetorial  $\times$  é um outro exemplo de álgebra não comutativa.

EXEMPLO 1.2. O espaço  $\mathcal{C}(\Omega)$  de funções reais contínuas em um domínio  $\Omega \subset \mathbb{R}^n$ ,  $n \in \mathbb{N}$ , é uma álgebra comutativa quando munida da multiplicação de funções  $(fg)(\mathbf{x}) = f(\mathbf{x})g(\mathbf{x})$ ,  $\forall \mathbf{x} \in \Omega$ , para  $f, g \in \mathcal{C}(\Omega)$ . Da mesma forma, o subespaço  $\mathcal{P}(\Omega)$  de polinômios restritos a  $\Omega$  também é uma álgebra associativa, sendo, portanto, uma *sub-álgebra* de  $\mathcal{C}(\Omega)$ . Ambas são álgebras com unidade para a multiplicação, dada pela função constante  $1(\mathbf{x}) = 1$ .

EXEMPLO 1.3. Álgebras aparecem de forma importante em diversos resultados. Por exemplo, o Teorema de Weierstrass (que diz que toda função contínua em um intervalo fechado limitado pode ser aproximada uniformemente por polinômios) tem por extensão o Teorema de Stone-Weierstrass [4]. Esse teorema trata da densidade na álgebra  $\mathcal{C}(K)$  das funções contínuas em um espaço métrico compacto  $K$  qualquer e diz que toda sub-álgebra de  $\mathcal{C}(K)$  que contém a unidade multiplicativa e que separa os pontos de  $K$  é densa em  $\mathcal{C}(K)$ . A condição de separar pontos é a de que, dados pontos  $x, y \in K$ , existe  $p$  na sub-álgebra tal que  $p(x) \neq p(y)$ .

OBSERVAÇÃO 1.3. O Exemplo 1.3 ilustra um caso em que um subespaço vetorial próprio é denso no espaço. Isso só é permitido acontecer em dimensão infinita. Em relação a isso, veja Observação 2.5.

EXEMPLO 1.4 (Transformada de Fourier no espaço de Schwartz). A transformada de Fourier definida no Exemplo 1.12 do Capítulo 3 pode ser considerada como um operador linear  $F : \mathcal{S}(\mathbb{R}^n, \mathbb{C}) \rightarrow \mathcal{S}(\mathbb{R}^n, \mathbb{C})$ , no espaço de Schwartz

$$\mathcal{S}(\mathbb{R}^n, \mathbb{C}) = \left\{ f \in \mathcal{C}^\infty(\mathbb{R}^n, \mathbb{C}); \forall \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}^n, \sup_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{x}^\alpha (D^\beta f)(\mathbf{x})\|_{\mathbb{R}^n} < \infty \right\},$$

onde  $\mathbf{x}^\alpha = (x_1^{\alpha_1}, \dots, x_n^{\alpha_n})$  e  $D^\beta f = (\partial_{x_1}^{\beta_1} f, \dots, \partial_{x_n}^{\beta_n} f)$ , para  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$  e  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_n) \in \mathbb{N}^n$ , que é o espaço das funções teste com decaimento superpolinomial das funções e das suas derivadas.

EXEMPLO 1.5 (Transformada inversa de Fourier no espaço de Schwartz). A *transformada inversa de Fourier*,  $G : \mathcal{S}(\mathbb{R}^n, \mathbb{C}) \rightarrow \mathcal{S}(\mathbb{R}^n, \mathbb{C})$  tem a seguinte forma, bem

semelhante à da própria transformada direta,

$$\hat{g}(\mathbf{x}) = \mathbf{G}(g)(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} g(\boldsymbol{\omega}) e^{i\mathbf{x} \cdot \boldsymbol{\omega}} d\boldsymbol{\omega}, \quad \mathbf{x} \in \mathbb{R}^n. \quad (1.1)$$

EXEMPLO 1.6 (Transformada de Fourier em  $L^2$ ). Pela densidade de  $\mathcal{S}(\mathbb{R}^n, \mathbb{C})$  em  $L^2(\mathbb{R}^n, \mathbb{C})$ , a transformada de Fourier pode ser estendida, também, a um operador linear em  $L^2(\mathbb{R}^n, \mathbb{C})$ .

### 1.3. Operadores lineares invertíveis.

PROPOSIÇÃO 1.3. *Seja  $X$  um espaço vetorial e considere a álgebra de operadores lineares  $\mathcal{L}(X)$ . O subconjunto de  $\mathcal{L}(X)$  formado pelos operadores lineares invertíveis é um grupo, com o elemento unitário sendo o operador identidade, denotado por  $\mathbf{l} : X \rightarrow X$  e definido por  $\mathbf{l}(\mathbf{u}) = \mathbf{u}$ , para todo  $\mathbf{u} \in X$ . Por vezes, denotamos o operador identidade por  $\mathbf{l}_X$ , para deixar claro o espaço vetorial em que ele age. Esse grupo de operadores lineares invertíveis em  $X$  é denotado por  $\text{GL}(X)$  (advindo de grupo linear geral ou general linear group, em inglês). No caso particular em que  $X = \mathbb{K}^n$ , denotamos esse espaço por  $\text{GL}(n, \mathbb{K})$ .*

OBSERVAÇÃO 1.4. O exemplo da **Observação 1.2** serve para ilustrar, também, um par de operadores  $\mathbf{T}, \mathbf{S} : X \rightarrow X$  em que  $\mathbf{S} \circ \mathbf{T} = \mathbf{l}$  é o operador identidade mas  $\mathbf{T} \circ \mathbf{S} \neq \mathbf{l}$ , não. E de que eles não são operadores invertíveis em  $X$ . Em dimensão finita, no entanto, isso não acontece, conforme provado no **Teorema 1.1**.

OBSERVAÇÃO 1.5. A nomenclatura de grupo linear geral vem do fato de que, no caso de uma matriz  $n \times n$  invertível, as suas colunas (e as suas linhas) são vetores linearmente independentes, que é a posição *genérica* (ou mais provável, em um sentido topológico) de  $n$  vetores no espaço. Ou seja, se escolhermos “aleatoriamente”  $n$  vetores em um espaço de dimensão  $n$  é quase certo que eles sejam LI. Pensemos indutivamente: tendo escolhido  $k$  vetores LI e escolhendo o  $(k+1)$ -ésimo vetor “aleatoriamente”, o conjunto será LD se, e somente se, o  $(k+1)$ -ésimo vetor estiver no subespaço  $k$ -dimensional gerado pelos  $k$  primeiros vetores, que é um espaço “ínfimo” dentro de todo o espaço  $n$ -dimensional possível. Nesse sentido, vemos que essa escolha é muito particular e improvável. Vejamos, se  $S_k$  é o subespaço gerado pelos  $k$  primeiros vetores LI e  $S'_k$  é um subespaço complementar, então o  $(k+1)$ -ésimo vetor será escolhido da forma  $\mathbf{u} = \mathbf{v} + \mathbf{w}$ , onde  $\mathbf{v} \in S_k$  e  $\mathbf{w} \in S'_k$  são arbitrários, sendo que a única escolha possível para que o conjunto agregado de  $(k+1)$  vetores seja LD é que  $\mathbf{w} = \mathbf{0}$ , que é uma escolha muito particular, perto de todas as outras escolhas possíveis. Uma maneira rigorosa de expressar isso é que o conjunto  $S_n \setminus S_k$  é aberto e denso no conjunto  $S_n$  de todas as escolhas possíveis. Há uma definição mais formal

de *propriedade genérica* em espaços topológicos que é justamente baseada na ideia de interseções enumeráveis de conjuntos abertos densos.

**TEOREMA 1.1.** *Seja  $X$  um espaço vetorial de dimensão finita e sejam  $T, S \in \mathcal{L}(X)$  operadores lineares em  $X$ . Então  $S \circ T = I$  se, e somente se,  $T \circ S = I$  se, e somente se,  $T$  e  $S$  são invertíveis com  $S = T^{-1}$  e vice-versa.*

**DEMONSTRAÇÃO.** Observe que, por simetria, provar que  $S \circ T = I$  implica em  $T \circ S = I$  é análogo a mostrar que  $T \circ S = I$  implica em  $S \circ T = I$ . Ou seja, basta mostrar uma direção para provar a equivalência.

Além disso, se  $T$  e  $S$  são invertíveis com  $S = T^{-1}$  e  $T = S^{-1}$ , então é imediato que  $S \circ T = I$  e  $T \circ S = I$ .

Ou seja, só precisamos mostrar que  $S \circ T = I$  implica em  $T$  e  $S$  serem bijetivos e  $T \circ S = I$ .

Caso  $X = \{\mathbf{0}\}$  seja o espaço trivial, então  $T$  e  $S$  levam o elemento neutro no elemento neutro e não há nada mais a se provar. Só precisamos trabalhar no caso em que  $\dim(X) \in \mathbb{N}$ .

Seja, então,  $n = \dim(X)$ . Suponha que  $S \circ T = I$ . Então  $\ker(S \circ T) = \ker(I) = \{\mathbf{0}\}$ . Além disso,  $\ker(T) \subset \ker(S \circ T)$ , logo  $\ker(T) = \{\mathbf{0}\}$  também, o que prova que  $T$  é injetiva. Pelo **Teorema 2.1** do Núcleo e da Imagem,

$$\dim(\text{Im}(T)) = \dim(X) - \dim(\ker(T)) = \dim(X).$$

Como  $X$  é de dimensão finita, isso implica em  $\text{Im}(T) = X$ , o que prova que  $T$  é sobrejetiva. Ou seja,  $T$  é uma bijeção. Resta mostrar que  $T \circ S = I$ .

Seja, agora, então,  $\mathbf{v} \in X$  arbitrário. Como  $T$  é uma bijeção, existe  $\mathbf{u} \in X$  tal que

$$\mathbf{v} = T(\mathbf{u}).$$

Aplicando  $S$  e usando que  $S \circ T = I$ , obtemos

$$S(\mathbf{v}) = S(T(\mathbf{u})) = \mathbf{u}.$$

Aplicando agora  $T$ , obtemos

$$T(S(\mathbf{v})) = T(\mathbf{u}) = \mathbf{v}.$$

Como isso vale para  $\mathbf{v}$  arbitrário, obtemos que  $T \circ S = I$ . Isso completa a demonstração.  $\square$

**OBSERVAÇÃO 1.6.** Olhando a demonstração do **Teorema 1.1**, parte do argumento foi mostrar que se  $T$  é invertível e  $S \circ T = I$ , então necessariamente  $S = T^{-1}$  e  $T \circ S = I$ . Mostramos isso tomando um  $\mathbf{v} \in X$  e chegando a  $T(S(\mathbf{v})) = \mathbf{v}$ . Outra maneira passa por usar a associatividade da composição e chegar em

$$S = S \circ I = S \circ (T \circ T^{-1}) = (S \circ T) \circ T^{-1} = I \circ T^{-1} = T^{-1}.$$

Assim,  $\mathsf{T} \circ \mathsf{S} = \mathsf{T} \circ \mathsf{T}^{-1} = \mathsf{I}$ .

OBSERVAÇÃO 1.7. O resultado do [Teorema 1.1](#) não se estende a espaços de dimensão infinita. De fato, considere, por exemplo, o espaço vetorial  $X = \mathbb{R}^{\mathbb{N}} = \{(x_j)_{j \in \mathbb{N}}\}$  das seqüências de números reais e o operador linear em  $X$  dado por  $\mathsf{T}(x_1, x_2, \dots) = (0, x_1, x_2, \dots)$ . Então  $\mathsf{T}$  é injetivo mas não é sobrejetivo.

Como consequência do [Teorema 2.1](#), em particular do [Teorema 2.3](#), temos que um operador  $\mathsf{T} \in \mathcal{L}(X)$  é invertível se e somente se é injetivo e se, e somente se, é sobrejetivo.

TEOREMA 1.2. *Seja  $X$  um espaço vetorial de dimensão finita e seja  $\mathsf{T} \in \mathcal{L}(X)$ . Então  $\ker(\mathsf{T}) = \{\mathbf{0}\}$  se, e somente se,  $\text{Im}(\mathsf{T}) = Y$  se, e somente se,  $\mathsf{T}$  é invertível.*

**1.4. Mudança de base.** Dado um vetor  $\mathbf{u}$  em um espaço vetorial  $X$  de dimensão finita  $n = \dim(X) \in \mathbb{N}$  sobre um corpo  $\mathbb{K}$ , podemos representá-lo em diferentes bases. De acordo com o [Teorema 3.1](#), se  $\mathfrak{a}$  e  $\mathfrak{b}$  são duas bases de  $X$ , então vale

$$(\mathbf{u})_{\mathfrak{b}} = (\mathbf{lu})_{\mathfrak{b}} = [\mathbb{I}]_{\mathfrak{b}}^{\mathfrak{a}}(\mathbf{u})_{\mathfrak{a}}.$$

A representação  $[\mathbb{I}]_{\mathfrak{b}}^{\mathfrak{a}}$  do operador identidade, da base  $\mathfrak{a}$  para a base  $\mathfrak{b}$ , tem o nome especial de *matriz de mudança de base*.

DEFINIÇÃO 1.4. *Seja  $X$  um espaço vetorial de dimensão finita  $n = \dim(X) \in \mathbb{N}$  e sejam  $\mathfrak{a}$  e  $\mathfrak{b}$  duas bases de  $X$ . Então a representação  $[\mathbb{I}]_{\mathfrak{b}}^{\mathfrak{a}}$  do operador identidade da base  $\mathfrak{a}$  para a base  $\mathfrak{b}$  é chamada de **matriz mudança de base**, da base  $\mathfrak{a}$  para a base  $\mathfrak{b}$ , e leva a representação de um vetor na base  $\mathfrak{a}$  na representação do vetor na base  $\mathfrak{b}$ , i.e.*

$$[\mathbf{u}]_{\mathfrak{b}} = [\mathbb{I}]_{\mathfrak{b}}^{\mathfrak{a}}[\mathbf{u}]_{\mathfrak{a}}.$$

TEOREMA 1.3. *No contexto da [Definição 1.4](#), a matriz  $[\mathbb{I}]_{\mathfrak{a}}^{\mathfrak{b}}$  define a operação inversa e vale a identidade*

$$[\mathbb{I}]_{\mathfrak{a}}^{\mathfrak{b}}[\mathbb{I}]_{\mathfrak{b}}^{\mathfrak{a}} = [\mathbb{I}]_{\mathfrak{a}} = \mathbf{I}.$$

DEMONSTRAÇÃO. Segue diretamente de

$$(\mathbf{u})_{\mathfrak{a}} = (\mathbf{lu})_{\mathfrak{a}} = [\mathbb{I}]_{\mathfrak{a}}^{\mathfrak{b}}(\mathbf{u})_{\mathfrak{b}} = [\mathbb{I}]_{\mathfrak{a}}^{\mathfrak{b}}(\mathbf{lu})_{\mathfrak{b}} = [\mathbb{I}]_{\mathfrak{a}}^{\mathfrak{b}}[\mathbb{I}]_{\mathfrak{b}}^{\mathfrak{a}}(\mathbf{u})_{\mathfrak{a}}.$$

ou seja,

$$[\mathbb{I}]_{\mathfrak{a}}^{\mathfrak{b}}[\mathbb{I}]_{\mathfrak{b}}^{\mathfrak{a}} = \mathbf{I}.$$

□

OBSERVAÇÃO 1.8. Seja  $X$  um espaço vetorial de dimensão finita  $n = \dim(X) \in \mathbb{N}$  e sejam  $\mathcal{a} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  e  $\mathcal{b} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  duas bases de  $X$ , com as suas respectivas bases duais  $\mathcal{a}^* = \{\mathbf{f}_1, \dots, \mathbf{f}_n\}$  e  $\mathcal{b}^* = \{\mathbf{g}_1, \dots, \mathbf{g}_n\}$ . Então, de acordo com o Teorema 3.1, seguindo a Definição 3.5, temos

$$[\mathbb{I}]_{\mathcal{b}}^{\mathcal{a}} = \begin{bmatrix} \langle \mathbf{g}_1, \mathbf{w}_1 \rangle & \langle \mathbf{g}_1, \mathbf{w}_2 \rangle & \cdots & \langle \mathbf{g}_1, \mathbf{w}_n \rangle \\ \langle \mathbf{g}_2, \mathbf{w}_1 \rangle & \langle \mathbf{g}_2, \mathbf{w}_2 \rangle & \cdots & \langle \mathbf{g}_2, \mathbf{w}_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{g}_m, \mathbf{w}_1 \rangle & \langle \mathbf{g}_m, \mathbf{w}_2 \rangle & \cdots & \langle \mathbf{g}_m, \mathbf{w}_n \rangle \end{bmatrix}$$

Em outras palavras, cada coluna  $j$  é a representação, na base  $\mathcal{b}$ , do  $j$ -ésimo vetor  $\mathbf{w}_j$  da base  $\mathcal{a}$ .

EXEMPLO 1.7. Considere a base canônica  $e = \{(1, 0), (0, 1)\}$  em  $\mathbb{R}^2$  e a base  $\mathcal{b} = \{(1, 2), (1, 3)\}$ . Temos

$$(1, 2) = 1(1, 0) + 2(0, 1), \quad (1, 3) = 1(1, 0) + 3(0, 1),$$

ou seja,

$$[\mathbb{I}]_e^{\mathcal{b}} = \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix}$$

Como estamos “chegando” na base canônica, as colunas são simplesmente os vetores da base  $\mathcal{b}$ . Para a mudança inversa,

$$(1, 0) = 3(1, 2) - 2(1, 3), \quad (0, 1) = -(1, 2) + (1, 3),$$

ou seja,

$$[\mathbb{I}]_{\mathcal{b}}^e = \begin{bmatrix} 3 & -1 \\ -2 & 1 \end{bmatrix}.$$

Para o controle de sanidade, verificamos os produtos das duas matrizes  $\mathbf{M} = [\mathbb{I}]_{\mathcal{b}}^e$  e  $\mathbf{M}^{-1} = [\mathbb{I}]_e^{\mathcal{b}}$ ,

$$\mathbf{M}^{-1}\mathbf{M} = [\mathbb{I}]_e^{\mathcal{b}}[\mathbb{I}]_{\mathcal{b}}^e = \begin{bmatrix} 3 & -1 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}.$$

Perfeito.

## 2. Similaridade

Operadores invertíveis formam uma classe bem particular de operadores lineares. E são fundamentais no estudo de outros operadores lineares, invertíveis ou não. Podemos pensar os operadores invertíveis como “reorganizações” de um mesmo espaço vetorial. Caso o espaço seja de dimensão finita, possuindo portanto uma base finita, digamos  $\mathcal{a} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ , um operador invertível  $\mathbf{M}$  funciona como uma mudança

de base, levando a base  $\mathcal{a}$  em uma nova base  $\mathcal{b} = \{M(\mathbf{w}_1), \dots, M(\mathbf{w}_n)\}$ . E o que acontece caso queiramos representar um determinado operador  $T \in \mathcal{L}(X)$  em cada uma dessas bases? Eles naturalmente devem ter representações (matriciais) diferentes em cada base, mas, no fundo, representam a mesma transformação. Como podemos identificar que representam a mesma transformação? Para isso, deve ser indiferente aplicar a transformação antes ou depois da mudança de base. Isso nos leva ao conceito de similaridade. E como operadores lineares em dimensão infinita estão intrinsecamente ligados a matrizes quadradas, o conceito de similaridade se estende a essas matrizes.

### 2.1. Operadores similares.

DEFINIÇÃO 2.1. *Seja  $X$  um espaço vetorial. Dois operadores lineares  $S, T \in \mathcal{L}(X)$  são ditos **similares**, ou **conjugados** quando existe um operador invertível  $M \in GL(X)$  tal que*

$$MS = TM,$$

ou, equivalentemente,

$$S = M^{-1}TM.$$

DEFINIÇÃO 2.2. *Seja  $X$  um espaço vetorial. Dado  $M \in GL(X)$ , a transformação  $T \mapsto T_M = M^{-1}TM$  em  $\mathcal{L}(X)$  é chamada de **transformação de similaridade** ou **operador de similaridade** ou **operador de conjugação**.*

TEOREMA 2.1. *Seja  $X$  um espaço vetorial sobre um corpo  $\mathbb{K}$ . O operador de similaridade satisfaz*

$$(S + T)_M = S_M + T_M,$$

$$(\lambda T)_M = \lambda T_M$$

e

$$(ST)_M = S_M T_M,$$

para quaisquer  $T, S \in \mathcal{L}(X)$  e  $\lambda \in \mathbb{K}$  e para todo  $M \in GL(X)$ .

DEMONSTRAÇÃO. Segue imediatamente da definição de similaridade e das propriedades de associatividade e de distributividade à esquerda e à direita da operação de composição que

$$(S + T)_M = M^{-1}(S + T)M = M^{-1}(SM + TM) = M^{-1}SM + M^{-1}TM = S_M + T_M$$

e

$$(\lambda T)_M = M^{-1}(\lambda T)M = M^{-1}(\lambda(TM)) = \lambda M^{-1}TM = \lambda T_M.$$

Usando ainda que  $I = MM^{-1}$ , temos, também, que

$$\begin{aligned} (\mathcal{S}\mathcal{T})_{\mathbf{M}} &= \mathbf{M}^{-1}(\mathcal{S}\mathcal{T})\mathbf{M} = \mathbf{M}^{-1}(\mathcal{S}\mathcal{I}\mathcal{T})\mathbf{M} = \mathbf{M}^{-1}(\mathcal{S}(\mathbf{M}\mathbf{M}^{-1})\mathcal{T})\mathbf{M} \\ &= (\mathbf{M}^{-1}\mathcal{S}\mathbf{M})(\mathbf{M}^{-1}\mathcal{T}\mathbf{M}) = \mathcal{S}_{\mathbf{M}}\mathcal{T}_{\mathbf{M}}. \end{aligned}$$

□

OBSERVAÇÃO 2.1. Em outras palavras, o **Teorema 2.1** diz que cada operador de similaridade  $\mathcal{T} \mapsto \mathcal{T}_{\mathbf{M}}$  é um *automorfismo* da álgebra  $\mathcal{L}(X)$ , pois preserva a sua estrutura de álgebra, ou seja, a de espaço vetorial munido da operação de composição.

TEOREMA 2.2. *Seja  $X$  um espaço vetorial. Dados um operador linear  $\mathcal{T} \in \mathcal{L}(X)$  e operadores lineares invertíveis  $\mathbf{M}, \mathbf{N} \in \text{GL}(X)$ , temos*

$$(\mathcal{T}_{\mathbf{M}})_{\mathbf{N}} = \mathcal{T}_{\mathbf{MN}}.$$

DEMONSTRAÇÃO. O resultado segue da associatividade da composição e do fato de que  $(\mathbf{MN})^{-1} = (\mathbf{N}^{-1}\mathbf{M}^{-1})$ , visto que

$$\begin{aligned} (\mathcal{T}_{\mathbf{M}})_{\mathbf{N}} &= (\mathbf{M}^{-1}\mathcal{T}\mathbf{M})_{\mathbf{N}} = \mathbf{N}^{-1}(\mathbf{M}^{-1}\mathcal{T}\mathbf{M})\mathbf{N} = (\mathbf{N}^{-1}\mathbf{M}^{-1})\mathcal{T}(\mathbf{MN}) \\ &= (\mathbf{MN})^{-1}\mathcal{T}(\mathbf{MN}) = (\mathcal{T})_{\mathbf{MN}}. \end{aligned}$$

□

EXEMPLO 2.1. O **Teorema 2.2** diz que o conjunto de operadores de similaridade  $\{\mathcal{S}(\mathbf{M}); \mathbf{M} \in \text{GL}(X)\}$ , onde  $\mathcal{S}(\mathbf{M}) \in \mathcal{L}(X)$  é dado por  $\mathcal{S}(\mathbf{M})(\mathcal{T}) = \mathcal{T}_{\mathbf{M}}$ , é um *grupo*, visto que  $\mathcal{S}(\mathbf{M})\mathcal{S}(\mathbf{N}) = \mathcal{S}(\mathbf{MN})$ , com o elemento neutro sendo  $\mathcal{S}(I)$  e o inverso de  $\mathcal{S}(\mathbf{M})$  sendo  $\mathcal{S}(\mathbf{M}^{-1})$ . Mas isso é mera curiosidade. Não iremos usar essa abstração aqui.

**2.2. Similaridade como relação de equivalência.** Os resultados acima garantem que a relação de similaridade entre operadores lineares é uma relação de equivalência.

TEOREMA 2.3. *Seja  $X$  um espaço vetorial. A relação de similaridade entre dois operadores  $\mathcal{T}, \mathcal{S} \in \mathcal{L}(X)$ , i.e.  $\mathcal{T} \sim \mathcal{S}$  quando  $\mathcal{S} = \mathcal{T}_{\mathbf{M}}$  para algum  $\mathbf{M} \in \text{GL}(X)$ , é uma relação de equivalência em  $\mathcal{L}(X)$ , ou seja, satisfaz*

- (i) (reflexividade)  $\mathcal{T} \sim \mathcal{T}$ ;
- (i) (simetria)  $\mathcal{T} \sim \mathcal{S}$  se, e somente se,  $\mathcal{S} \sim \mathcal{T}$ .
- (i) (transitividade)  $\mathcal{T} \sim \mathcal{S}$  e  $\mathcal{S} \sim \mathcal{R}$ , então  $\mathcal{T} \sim \mathcal{R}$ .

DEMONSTRAÇÃO. Para a reflexividade, basta considerar o operador identidade  $I \in \text{GL}(X)$ , de tal modo que  $\mathcal{T}_I = I^{-1}\mathcal{T}I = I\mathcal{T} = \mathcal{T}$ , ou seja,  $\mathcal{T} \sim \mathcal{T}$ , para um  $\mathcal{T} \in \mathcal{L}(X)$  arbitrário.

Para a simetria, se  $T \sim S$ , dado  $M \in GL(X)$  tal que  $S = M^{-1}TM$ , basta considerar  $M^{-1}$  no operador de similaridade, para obter  $T = MSM^{-1} = (M^{-1})^{-1}SM^{-1} = S_{M^{-1}}$ , mostrando que  $S \sim T$ .

Por fim, se  $T \sim S$  e  $S \sim R$ , então existem  $M, N \in GL(X)$  tais que  $S = M^{-1}TM$  e  $R = N^{-1}TN$ . Com isso,

$$R = N^{-1}TN = N^{-1}(M^{-1}TM)N = (N^{-1}M^{-1})T(MN) = (MN)^{-1}T(MN) = T_{MN},$$

ou seja,  $T \sim R$ , completando a demonstração. Observe que, para a transitividade, usamos a propriedade provada no [Teorema 2.2](#).  $\square$

O resultado do [Teorema 2.3](#) de que a similaridade é uma relação de equivalência nos leva a buscar entender o conjunto de todos os operadores lineares a partir de representantes de cada classe de equivalência

$$[[T]]_{\sim} = \{T_M; \forall M \in GL(X)\}$$

em  $\mathcal{L}(X)/\sim$ , e, assim, buscar, também, classificar todos os tipos de operadores possíveis. Veremos que isso é possível em dimensão finita.

Algumas classes são simples, como mostram os seguintes exemplos.

**EXEMPLO 2.2.** A classe de equivalência do operador identidade  $I$  contém apenas o operador identidade, i.e.  $[[I]]_{\sim} = \{I\}$ . De fato, se  $M \in GL(X)$ , então  $I_M = M^{-1}IM = M^{-1}M = I$ . Da mesma forma, a classe de equivalência do operador nulo  $0$  (definido por  $0(\mathbf{u}) = \mathbf{0} \in X$ , para todo  $\mathbf{u} \in X$ ) contém apenas o operador nulo, i.e.  $[[0]]_{\sim} = \{0\}$ , visto que, nesse caso,  $0_M = M^{-1}0M = 0$ .

**EXEMPLO 2.3.** Mais geralmente, a classe de equivalência de um múltiplo da identidade,  $\lambda I$ , para  $\lambda \in \mathbb{K}$ , que inclui os casos anteriores  $\lambda = 1$  e  $\lambda = 0$ , também contém apenas o próprio operador, i.e.  $[[\lambda I]]_{\sim} = \{\lambda I\}$ , visto que, se  $M \in GL(X)$ , então  $(\lambda I)_M = M^{-1}(\lambda I)M = \lambda M^{-1}IM = \lambda M^{-1}M = \lambda I$ .

Vimos que o grupo de operadores lineares não é comutativo, mas há uma relação interessante entre as duas ordens possíveis de se compor dois operadores, como mostra o resultado a seguir.

**TEOREMA 2.4.** *Seja  $X$  um espaço vetorial e sejam  $T, S \in \mathcal{L}(X)$ . Se pelo menos um dos dois operadores for invertível, então  $TS$  e  $ST$  são similares entre si.*

**DEMONSTRAÇÃO.** Supondo  $T$  invertível, usamos  $M = T$  no operador de similaridade para escrever

$$ST = T^{-1}TST = T^{-1}(TS)T = (TS)_T,$$

mostrando que  $ST$  é similar a  $TS$ .

Por simetria, o resultado também vale caso  $S$  seja invertível. Ou, mais explicitamente, supondo  $S$  invertível, temos

$$TS = S^{-1}STS = (ST)_S,$$

completando a demonstração.  $\square$

EXEMPLO 2.4. Em relação ao resultado do Teorema 2.4, definindo  $M = T$ , podemos, naturalmente, escrever

$$MST = TST = TSM.$$

Ou seja, para que isso implique em similaridade, precisamos que  $M = T$  seja invertível. Da mesma forma, escolhendo  $M = S$ , podemos escrever

$$MTS = STS = STM.$$

Nesse caso, precisamos que  $M = S$  seja invertível para que isso implique em similaridade.

EXEMPLO 2.5. Vimos na Observação 1.1, que a rotação de  $90^\circ$  no sentido trigonométrico  $T(x, y) = (-y, x)$  e a reflexão  $S(x, y) = (x, -y)$  no eixo  $x$  não comutam, como operadores em  $\mathbb{R}^2$ . Mas observe que ambos são invertíveis. Então segue do Teorema 2.4 que  $ST$  é similar a  $TS$ , com a similaridade podendo ser obtida tanto com  $T$  quanto com  $S$ . Vimos que  $(ST)(x, y) = (-y, -x)$  é a reflexão em relação à reta  $y = -x$ , enquanto que  $(TS)(x, y) = (y, x)$  é a reflexão em relação à reta  $y = x$ . Considerando, por exemplo,  $M = T$  no operador de similaridade, temos

$$MST = TSM,$$

com

$$(MST)(x, y) = T((ST)(x, y)) = T(-y, -x) = (x, -y)$$

Considerando  $M = S$  no operador de similaridade, temos

$$MTS = STM,$$

com

$$(MTS)(x, y) = S((TS)(x, y)) = S(y, x) = (y, -x).$$

**2.3. Similaridade entre matrizes.** Matrizes são, de certa forma, a essência das transformações lineares em dimensão finita. Já vimos que toda matriz dá origem a uma transformação linear entre espaços  $\mathbb{R}^n$  e  $\mathbb{R}^m$ , para  $n, m \in \mathbb{N}$  apropriados, e toda transformação linear entre espaços de dimensão finita pode ser representada por matrizes, dependendo das bases escolhidas para cada espaço.

No caso de operadores lineares, o domínio e o contradomínio são os mesmos e a representação de um operador linear em uma determinada base dá origem a uma

matriz quadrada,  $n \times n$ , onde  $n$  é a dimensão do espaço. Escolhendo bases diferentes, obtemos matrizes diferentes representando a mesma transformação linear. Nesse sentido, o conceito de similaridade se aplica, identificando que essas matrizes são manifestações de um mesmo operador linear. Vamos ver, aqui, mais diretamente, o efeito desse conceito na espaço das matrizes.

Se  $T \in \mathcal{L}(X)$  é uma transformação linear em um espaço  $X$  de dimensão finita  $n = \dim(X) \in \mathbb{N}$  sobre um corpo  $\mathbb{K}$  e se  $\alpha$  e  $\beta$  são duas bases de  $X$ , então temos as representações  $[T]_\alpha$  e  $[T]_\beta$  como matrizes em  $\mathbb{K}^n$ . Qual a relação entre  $[T]_\alpha$  e  $[T]_\beta$ ? Isso é dado pelo próximo resultado.

**TEOREMA 2.5.** *Seja  $T \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  de dimensão finita  $n = \dim(X) \in \mathbb{N}$  e sejam  $\alpha$  e  $\beta$  duas bases de  $X$ . Então*

$$[T]_\beta = [I]_\beta^\alpha [T]_\alpha [I]_\alpha^\beta \quad (2.1)$$

**DEMONSTRAÇÃO.** Certamente podemos escrever

$$(T\mathbf{u})_\beta = [T]_\beta(\mathbf{u})_\beta, \quad (T\mathbf{u})_\alpha = [T]_\alpha(\mathbf{u})_\alpha,$$

assim como

$$(T\mathbf{u})_\beta = [I]_\beta^\alpha (T\mathbf{u})_\alpha = [I]_\beta^\alpha [T]_\alpha(\mathbf{u})_\alpha = [I]_\beta^\alpha [T]_\alpha [I]_\alpha^\beta(\mathbf{u})_\beta.$$

Combinando as expressões, temos

$$[T]_\beta(\mathbf{u})_\beta = [I]_\beta^\alpha [T]_\alpha [I]_\alpha^\beta(\mathbf{u})_\beta,$$

para todo  $\mathbf{u}$ . Ou seja,

$$[T]_\beta = [I]_\beta^\alpha [T]_\alpha [I]_\alpha^\beta.$$

□

Caso o espaço seja o próprio  $\mathbb{K}^n$  e estejamos trabalhando diretamente com matrizes como os operadores lineares, a identidade (2.1) toma a forma

$$\mathbf{B} = \mathbf{M}^{-1}\mathbf{A}\mathbf{M},$$

onde  $\mathbf{M}$  faz o papel da matriz mudança de base  $[I]_\alpha^\beta$ , com inversa  $\mathbf{M}^{-1} = [I]_\beta^\alpha$ ;  $\mathbf{A}$  faz o papel da representação do operador linear na base  $\alpha$  e  $\mathbf{B}$ , na base  $\beta$ . Isso nos leva à seguinte definição.

**DEFINIÇÃO 2.3.** *Duas matrizes  $\mathbf{A}, \mathbf{B} \in \mathbb{K}^{n \times n}$ ,  $n \in \mathbb{N}$ , são ditas **similares** quando existe uma matriz invertível  $\mathbf{M} \in \mathbb{K}^{n \times n}$  tal que*

$$\mathbf{M}\mathbf{B} = \mathbf{A}\mathbf{M},$$

ou, equivalentemente,

$$\mathbf{B} = \mathbf{M}^{-1}\mathbf{A}\mathbf{M}.$$

EXEMPLO 2.6. As matrizes

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -6 & 5 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

são similares, i.e.  $\mathbf{B} = \mathbf{M}^{-1}\mathbf{A}\mathbf{M}$ , onde

$$\mathbf{M} = \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix}, \quad \mathbf{M}^{-1} = \begin{bmatrix} 3 & -1 \\ -2 & 1 \end{bmatrix},$$

ou seja,

$$\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 3 & -1 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -6 & 5 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix}.$$

A matriz  $\mathbf{A}$  representa o operador

$$\mathbb{T}(x, y) = (y, y - 6x)$$

na base canônica, enquanto que  $\mathbf{B}$  representa o operador na base formada pelas colunas de  $\mathbf{M}$ , i.e.

$$\mathcal{B} = \{(1, 2), (1, 3)\}.$$

A matriz  $\mathbf{M} = [\mathbb{I}]_{\mathcal{B}}^{\mathcal{E}}$  é a matriz mudança de base, da base canônica para a base  $\mathcal{B}$ , com matriz inversa  $\mathbf{M}^{-1} = [\mathbb{I}]_{\mathcal{E}}^{\mathcal{B}}$ .

### 3. Determinante de matrizes e operadores linear

Vimos, na [Seção 4.1](#), o determinante como um funcional multilinear em  $(\mathbb{K}^n)^n = \mathbb{K}^n \times \cdots \times \mathbb{K}^n$ , representando o  $n$ -volume, com sinal, da região delimitada por  $n$  vetores. Esse mesmo determinante pode ser usado para indicar a taxa de variação de  $n$ -volume de uma região sendo levada em outra região por uma transformação linear. Isso nos leva ao conceito de determinante de uma matriz e, eventualmente, do de um operador linear, como veremos agora.

#### 3.1. Determinante de matrizes quadradas.

DEFINIÇÃO 3.1 (Determinante de matrizes quadradas e fórmula de Leibniz). *Seja  $n \in \mathbb{N}$  e  $\mathbb{K}$  um corpo e considere uma matriz  $\mathbf{A} \in \mathbb{K}^{n \times n}$ . Definimos o determinante de  $\mathbf{A}$  por*

$$\det \mathbf{A} = \det(\mathbf{A}\mathbf{e}_1, \dots, \mathbf{A}\mathbf{e}_n),$$

onde o determinante no lado direito é o funcional multilinear alternante e normalizado na base canônica, definido no [Teorema 4.2](#). Em termos dos coeficientes  $\mathbf{A} = (a_{ij})_{i,j=1}^n$  de  $\mathbf{A}$ , temos, também, a fórmula de Leibniz (4.8), dada por

$$\det(\mathbf{A}) = \sum_{\sigma \in S_n} (-1)^{N(\sigma)} a_{\sigma(1),1} a_{\sigma(2),2} \cdots a_{\sigma(n),n}, \quad (3.1)$$

OBSERVAÇÃO 3.1 (Notação). Uma notação comum, para simplificar, do determinante de  $n$  vetores  $\mathbf{u}_j = (a_{ij})_{i=1}^n$ , ou da matriz associada  $\mathbf{A} = (a_{ij})_{i,j=1}^{i,j=n}$ , é com barras simples para delimitar os coeficiente, ou seja

$$\det(\mathbf{u}_1, \dots, \mathbf{u}_n) = \det(\mathbf{A}) = \det \left( \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \right) = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}.$$

EXEMPLO 3.1. Em dimensão dois, com dois vetores  $\mathbf{u} = (a, c)$  e  $\mathbf{v} = (b, d)$ , formando a matriz

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

temos

$$\det(\mathbf{u}, \mathbf{v}) = \det(\mathbf{A}) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ac - bd.$$

Em dimensão três, com três vetores  $\mathbf{u} = (a, d, g)$ ,  $\mathbf{v} = (b, e, h)$ ,  $\mathbf{w} = (c, f, i)$ , formando a matriz

$$\mathbf{A} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix},$$

temos

$$\det(\mathbf{u}, \mathbf{v}, \mathbf{w}) = \det(\mathbf{A}) = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei - afh + bdi - bcg + cdh - ceg.$$

OBSERVAÇÃO 3.2 (Fórmula de Laplace). Há outras maneiras de se calcular o determinante. Por exemplo, pensando o determinante como uma função matricial, conforme discutido [Definição 3.1](#), a fórmula de Laplace, ou expansão de Laplace, calcula o determinante de uma matriz recursivamente em termos de determinantes de “matrizes menores”,

$$\det(\mathbf{A}) = \sum_{j=1}^n (-1)^{i+j} a_{ij} M_{ij},$$

em termos de qualquer linha  $i = 1, \dots, n$ , ou

$$\det(\mathbf{A}) = \sum_{i=1}^n (-1)^{i+j} a_{ij} M_{ij},$$

em termos de qualquer coluna  $j = 1, \dots, n$ , onde  $M_{ij}$  é a *matriz menor* obtida de  $\mathbf{A}$  descartando a linha  $i$  e a coluna  $j$ . Por exemplo, no caso de uma matriz  $3 \times 3$ , temos

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix}.$$

OBSERVAÇÃO 3.3 (Determinante via escalonamento). Conforme discutido na [Seção 3.5](#), o determinante pode ser calculado mais facilmente via escalonamento da matriz.

**3.2. Determinantes de matrizes especiais.** Podemos calcular facilmente o determinante de algumas matrizes especiais.

OBSERVAÇÃO 3.4 (Determinante da identidade). O determinante da matriz identidade é exatamente 1:

$$\det(\mathbf{I}) = \det(\mathbf{Ie}_1, \dots, \mathbf{Ie}_n) = \det(\mathbf{e}_1, \dots, \mathbf{e}_n) = 1.$$

OBSERVAÇÃO 3.5 (Determinante de uma matriz triangular superior). O determinante de uma matriz triangular superior é o produto das diagonais. De fato, se  $\mathbf{U} = (a_{ij})_{ij}$  com  $a_{ij} = 0$  para  $i > j$ , então, na fórmula (3.1), como os coeficientes do somatório são  $a_{\sigma(1),1} a_{\sigma(2),2} \cdots a_{\sigma(n),n}$ , o único termo que sobrevive deve ter  $\sigma(j) \leq j$ , para todo  $j = 1, \dots, n$ , o que necessariamente nos dá  $\sigma(1) = 1, \sigma(2) = 2, \dots, \sigma(n) = n$ , ou seja,  $\sigma$  é a identidade, nos dando  $N(\sigma) = 0$  e

$$\det(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n) = \sum_{\sigma \in S_n} (-1)^{N(\sigma)} a_{\sigma(1),1} a_{\sigma(2),2} \cdots a_{\sigma(n),n} = a_{11} \cdots a_{nn},$$

que é o produto dos termos da diagonal.

OBSERVAÇÃO 3.6 (Determinante de uma matriz triangular inferior). Da mesma forma, o determinante de uma matriz triangular inferior também é o produto dos elementos da diagonal.

OBSERVAÇÃO 3.7 (Determinante de uma matriz diagonal). Como caso particular, o determinante de uma matriz diagonal também é o produto dos elementos da diagonal. Isso também pode ser visto diretamente da fórmula de Leibniz (3.1).

OBSERVAÇÃO 3.8 (Determinante da matriz elementar de troca de linhas). Se  $\mathbf{E}$  é uma matriz elementar de troca de linhas, então  $\det(\mathbf{E}) = -1$ , pois equivale a trocar a posição dos elementos de uma base, ou seja,

$$\begin{aligned} \det(\mathbf{E}) &= \det(\dots, \mathbf{Ee}_i, \dots, \mathbf{Ee}_j, \dots) = \det(\dots, \mathbf{e}_j, \dots, \mathbf{e}_i, \dots) \\ &= -\det(\dots, \mathbf{e}_i, \dots, \mathbf{e}_j, \dots) = -1. \end{aligned}$$

para determinados índices  $1 \leq i < j \leq n$ .

**OBSERVAÇÃO 3.9** (Determinante da matriz elementar de multiplicação de uma linha por um escalar). Se  $\mathbf{E}$  é uma matriz elementar associada à multiplicação de uma linha por um escalar, então ela é da forma diagonal, com um fator  $\lambda$  na linha a ser multiplicada e com 1 nos outros elementos da diagonal. Dessa forma, o determinante de  $\mathbf{E}$ , sendo o produto dos elementos da diagonal, é igual a  $\det \mathbf{E} = \lambda$ .

**OBSERVAÇÃO 3.10** (Determinante de uma matriz elementar que adiciona um múltiplo de uma linha a outra). Nesse caso, se o efeito  $\mathbf{E}\mathbf{A}$  de uma matriz elementar  $\mathbf{E}$  em uma matriz  $\mathbf{A}$  é adicionar um múltiplo  $\lambda$  de uma linha  $i$  de  $\mathbf{A}$  à linha  $j$  de  $\mathbf{A}$ , então  $\mathbf{E}\mathbf{e}_l = \mathbf{e}_l$  para  $l \neq i$  e  $\mathbf{E}\mathbf{e}_i = \mathbf{e}_i + \lambda\mathbf{e}_j$  (verifique!). Com isso,

$$\begin{aligned} \det(\mathbf{E}) &= \det(\mathbf{E}\mathbf{e}_1, \dots, \mathbf{E}\mathbf{e}_i, \dots, \mathbf{E}\mathbf{e}_n) = \det(\mathbf{e}_1, \dots, \mathbf{e}_i + \lambda\mathbf{e}_j, \dots, \mathbf{e}_j, \dots, \mathbf{e}_n) \\ &= \det(\mathbf{e}_1, \dots, \mathbf{e}_i, \dots, \mathbf{e}_j, \dots, \mathbf{e}_n) + \lambda \det(\mathbf{e}_1, \dots, \mathbf{e}_j, \dots, \mathbf{e}_j, \dots, \mathbf{e}_n) = 1 + 0 = 1. \end{aligned}$$

visto que  $\mathbf{e}_j$  se repete, no último determinante, e, portanto, o segundo determinante se anula.

**3.3. Propriedades do determinante.** Definimos o determinante de  $n$  vetores em um espaço de dimensão  $n$  como um  $n$ -volume com sinal, inspirado na ideia de volume. É natural que, no caso de matrizes, o determinante esteja associado à taxa de variação de volume pela transformação associada à matriz.

**TEOREMA 3.1.** Se  $\mathbf{A} \in \mathbb{K}^n$ ,  $n \in \mathbb{N}$ , em um corpo  $\mathbb{K}$  e  $\mathbf{u}_1, \dots, \mathbf{u}_n \in \mathbb{K}^n$ , então

$$\det(\mathbf{A}\mathbf{u}_1, \dots, \mathbf{A}\mathbf{u}_n) = \det(\mathbf{A}) \det(\mathbf{u}_1, \dots, \mathbf{u}_n) \quad (3.2)$$

**DEMONSTRAÇÃO.** Escrevemos  $\mathbf{u}_j = (a_{1j}, \dots, a_{nj}) = a_{1j}\mathbf{e}_1 + \dots + a_{nj}\mathbf{e}_n$ . Assim,

$$\begin{aligned} \det(\mathbf{A}\mathbf{u}_1, \dots, \mathbf{A}\mathbf{u}_n) &= \det\left(\mathbf{A}\left(\sum_{j=1}^n a_{1j}\mathbf{e}_i\right), \dots, \mathbf{A}\left(\sum_{j=1}^n a_{nj}\mathbf{e}_i\right)\right) \\ &= \sum_{j_1=1}^n \cdots \sum_{j_n=1}^n a_{1j_1} \cdots a_{nj_n} \det(\mathbf{A}\mathbf{e}_{j_1}, \dots, \mathbf{A}\mathbf{e}_{j_n}) \\ &= \sum_{\sigma \in S_n} a_{1\sigma(1)} \cdots a_{n\sigma(n)} \det(\mathbf{A}\mathbf{e}_{\sigma(1)}, \dots, \mathbf{A}\mathbf{e}_{\sigma(n)}) \\ &= \sum_{\sigma \in S_n} (-1)^{N(\sigma)} a_{1\sigma(1)} \cdots a_{n\sigma(n)} \det(\mathbf{A}\mathbf{e}_1, \dots, \mathbf{A}\mathbf{e}_n). \end{aligned}$$

Pela **Definição 3.1** de  $\det(A)$  e pela fórmula (4.8) de  $\det(\mathbf{u}_1, \dots, \mathbf{u}_n)$ , isso nos dá exatamente (3.2).  $\square$

OBSERVAÇÃO 3.11. No contexto do Teorema 3.1, caso  $\mathbf{u}_1, \dots, \mathbf{u}_n$  sejam linearmente dependentes, então  $\mathbf{A}\mathbf{u}_1, \dots, \mathbf{A}\mathbf{u}_n$  também são linearmente dependentes e os dois lados de (3.2) se anulam. Caso esses vetores sejam linearmente independentes, então o determinante  $\det(\mathbf{u}_1, \dots, \mathbf{u}_n)$  é não nulo e podemos escrever o determinante de  $\mathbf{A}$  precisamente como taxa de variação de  $n$ -volume:

$$\det(\mathbf{A}) = \frac{\det(\mathbf{A}\mathbf{u}_1, \dots, \mathbf{A}\mathbf{u}_n)}{\det(\mathbf{u}_1, \dots, \mathbf{u}_n)}.$$

Pode ser que  $\mathbf{A}$  não seja invertível e que  $\mathbf{A}\mathbf{u}_1, \dots, \mathbf{A}\mathbf{u}_n$  sejam linearmente dependentes, caso este em que  $\det(\mathbf{A}) = 0$ . Mas, em geral, podemos ter  $\det(\mathbf{A})$  também diferente de zero, quantificando precisamente a variação de volume.

Uma consequência importante do resultado anterior diz respeito ao determinante do produto de matrizes.

TEOREMA 3.2. *Sejam  $\mathbf{A}, \mathbf{B} \in \mathbb{K}^n$ ,  $n \in \mathbb{N}$ , em um corpo  $\mathbb{K}$ . Então*

$$\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B}). \quad (3.3)$$

DEMONSTRAÇÃO. Segue diretamente da Definição 3.1 do determinante e da fórmula (3.2) que

$$\begin{aligned} \det(\mathbf{AB}) &= \det(\mathbf{A}\mathbf{B}\mathbf{e}_1, \dots, \mathbf{A}\mathbf{B}\mathbf{e}_n) \\ &= \det(\mathbf{A}) \det(\mathbf{B}\mathbf{e}_1, \dots, \mathbf{B}\mathbf{e}_n) \\ &= \det(\mathbf{A}) \det(\mathbf{B}), \end{aligned}$$

provando o resultado. □

Como  $\det(\mathbf{A}) \det(\mathbf{B}) = \det(\mathbf{B}) \det(\mathbf{A})$ , pois são escalares, temos o seguinte colôquio.

COROLÁRIO 3.1. *Dados  $\mathbf{A}, \mathbf{B} \in \mathbb{K}^n$ ,  $n \in \mathbb{N}$ , em um corpo  $\mathbb{K}$ , temos*

$$\det(\mathbf{AB}) = \det(\mathbf{BA}).$$

### 3.4. O determinante de matrizes invertíveis.

TEOREMA 3.3. *Uma matriz  $\mathbf{M} \in \mathbb{K}^{n \times n}$ ,  $n \in \mathbb{N}$ , sobre um corpo  $\mathbb{K}$ , é invertível se, e somente se,  $\det(\mathbf{M}) \neq 0$ .*

DEMONSTRAÇÃO. Pelo Teorema 2.1 do núcleo e da imagem, como  $\mathbf{M}$  é uma matriz quadrada, portanto gerando um operador linear de  $\mathbb{K}^n$  em  $\mathbb{K}^n$ , então  $\mathbf{M}$  é invertível se, e somente se,  $\text{Im}(\mathbf{M}) = \mathbb{K}^n$ , ou seja, se e somente se  $\{\mathbf{M}\mathbf{e}_1, \dots, \mathbf{M}\mathbf{e}_n\}$  gera  $\mathbb{K}^n$ , o que, como são  $n$  vetores, é equivalente a  $\{\mathbf{M}\mathbf{e}_1, \dots, \mathbf{M}\mathbf{e}_n\}$  ser linearmente independente. Logo, segue de Teorema 4.3, que  $\mathbf{M}$  é invertível se, e somente se,

$\det(\mathbf{M}\mathbf{e}_1, \dots, \mathbf{M}\mathbf{e}_n) \neq 0$ . Pela definição de  $\det(\mathbf{M})$ , isso significa que  $\mathbf{M}$  é invertível se, e somente se,  $\det(\mathbf{M}) \neq 0$ , completando a demonstração.  $\square$

Uma matriz não invertível é dita uma matriz singular, logo, temos o seguinte colorário.

**COROLÁRIO 3.2.** *Uma matriz  $\mathbf{A} \in \mathbb{K}^{n \times n}$ ,  $n \in \mathbb{N}$ , sobre um corpo  $\mathbb{K}$ , é singular se, e somente se,  $\det(\mathbf{A}) = 0$ .*

Para uma matriz invertível, temos uma relação imediata entre o determinante da matriz e a da sua inversa, graças à propriedade (3.3).

**TEOREMA 3.4.** *Seja  $\mathbf{M} \in \mathbb{K}^{n \times n}$  uma matriz invertível, sobre um corpo  $\mathbb{K}$ , com  $n \in \mathbb{N}$ . Então*

$$\det(\mathbf{M}^{-1}) = \frac{1}{\det(\mathbf{M})}, \quad (3.4)$$

com ambos os determinantes não nulos.

**DEMONSTRAÇÃO.** Como  $\mathbf{M}$  é invertível, o seu determinante é diferente de zero, assim como o de  $\mathbf{M}^{-1}$ . Além disso, graças à (3.3), temos

$$\det(\mathbf{I}) = \det(\mathbf{M}^{-1}\mathbf{M}) = \det(\mathbf{M}^{-1}) \det(\mathbf{M}).$$

Como  $\det(\mathbf{I}) = 1$ , obtemos o resultado desejado.  $\square$

**3.5. Calculando o determinante via escalonamento.** Combinando a fórmula (3.3) para o determinante do produto de matrizes com a fatorização LU vista na Seção 5.1, obtemos uma maneira mais prática de se calcular o determinante.

De fato, dada uma matriz  $\mathbf{A}$ , o escalonamento nos leva a uma fatorização

$$\mathbf{A} = \mathbf{L}\mathbf{U},$$

como em (5.3), onde  $\mathbf{U}$  é triangular superior e onde  $\mathbf{L}$  é uma composição

$$\mathbf{L} = \mathbf{E}_1^{-1}\mathbf{E}_2^{-1} \cdots \mathbf{E}_k^{-1}$$

de matrizes elementares  $\mathbf{E}_j^{-1}$  (lembrando que a inversa de uma matriz elementar também é elementar), conforme (5.2). Graças à fórmula (3.3), obtemos

$$\det(\mathbf{A}) = \det(\mathbf{E}_1^{-1}) \cdots \det(\mathbf{E}_k^{-1}) \det(\mathbf{U}).$$

Como  $\mathbf{U}$  é triangular, o seu determinante é o produto das diagonais. E o determinante de matrizes elementares também é facilmente calculável (veja Seção 3.2).

Segue, também, de (3.4), que  $\det(\mathbf{E}_j^{-1}) = 1/\det(\mathbf{E}_j)$ , com  $\det(\mathbf{E}_j)$  podendo ser deduzido mais diretamente do processo de escalonamento. Dessa forma, obtemos

$$\det(\mathbf{A}) = \frac{\det(\mathbf{U})}{\det(\mathbf{E}_1) \cdots \det(\mathbf{E}_k)}. \quad (3.5)$$

EXEMPLO 3.2. Incluir um exemplo de cálculo de determinante via escalonamento.

**3.6. Determinante invariante por similaridade.** Como consequência das fórmulas do determinante do produto de matrizes e do determinante da inversa de uma matriz, obtemos o resultado fundamental de que os determinantes de matrizes similares são iguais. Ou seja, o determinante é invariante pela operação de similaridade.

TEOREMA 3.5. *Sejam  $\mathbf{A}, \mathbf{B} \in \mathbb{K}^n$ ,  $n \in \mathbb{N}$ , sobre um corpo  $\mathbb{K}$ . Se  $\mathbf{A} \sim \mathbf{B}$  então  $\det(\mathbf{A}) = \det(\mathbf{B})$ .*

DEMONSTRAÇÃO. Com as matrizes sendo similares, existe uma matriz invertível  $\mathbf{M} \in \mathbb{K}^{n \times n}$  tal que  $\mathbf{B} = \mathbf{M}^{-1}\mathbf{A}\mathbf{M}$ . Com isso, segue de (3.3) que

$$\det(\mathbf{B}) = \det(\mathbf{M}^{-1}\mathbf{A}\mathbf{M}) = \det(\mathbf{M}^{-1}) \det(\mathbf{A}\mathbf{M}) = \det(\mathbf{M}^{-1}) \det(\mathbf{A}) \det(\mathbf{M}).$$

Graças a (3.4), obtemos  $\det(\mathbf{B}) = \det(\mathbf{A})$ , provando o resultado.  $\square$

**3.7. Determinante de operadores lineares.** Graças a invariância do determinante por similaridade, obtemos a importante propriedade do determinante da representação de um operador em uma base ser independente da base. Isso nos permite definir o determinante de um operador linear em um espaço vetorial qualquer de dimensão finita.

TEOREMA 3.6. *Seja  $\mathsf{T} \in \mathcal{L}(X)$ , em um espaço vetorial  $X$  de dimensão finita. Então  $\det([\mathsf{T}]_{\mathfrak{b}})$  independe da escolha da base  $\mathfrak{b}$  de  $X$ .*

DEMONSTRAÇÃO. Segue do Teorema 3.5 e do fato das representações de um operador em bases diferentes serem matrizes similares, conforme visto no Teorema 2.5.  $\square$

Com isso, podemos definir o determinante de um operador linear escolhendo uma base arbitrária.

DEFINIÇÃO 3.2. *Seja  $\mathsf{T} \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  de dimensão finita. Definimos o determinante de  $\mathsf{T}$  como sendo o determinante de uma representação qualquer de  $\mathsf{T}$ , ou seja,  $\det(\mathsf{T}) = \det([\mathsf{T}]_{\mathfrak{b}})$ , para uma base  $\mathfrak{b}$  arbitrária de  $X$ .*

OBSERVAÇÃO 3.12. Vale  $\det(\mathbb{T}) = d_{\mathfrak{b}}(\mathbb{T}\mathbf{u}_1, \dots, \mathbb{T}\mathbf{u}_n) / d_{\mathfrak{b}}(\mathbf{u}_1, \dots, \mathbf{u}_n)$ , onde  $d_{\mathfrak{b}} : X^n \rightarrow \mathbb{K}$  é qualquer funcional multilinear alternante não nulo em  $X^n$ , ou seja, essa razão é independente da escolha da base (veja [Observação 4.1](#) e [Exercício 5.41](#)).

O [Teorema 3.2](#) tem a sua versão de operadores.

TEOREMA 3.7. *Sejam  $\mathbb{T}, \mathbb{S} \in \mathcal{L}(X)$  operadores em um espaço vetorial  $X$  real ou complexo. Então*

$$\det(\mathbb{TS}) = \det(\mathbb{ST}) = \det(\mathbb{T}) \det(\mathbb{S}). \quad (3.6)$$

OBSERVAÇÃO 3.13. O resultado do [Teorema 3.7](#) é natural quando pensamos o determinante como taxa de variação de hipervolume, de modo que a taxa de variação pela composição  $\mathbb{TS}$  é o produto das taxas de variação por cada operador. E apesar do conjunto em si não ser o mesmo, dependendo da ordem dos operadores, a taxa de variação é a mesma.

**3.8. Determinante do operador adjunto.** Usando que a representação matricial do operador adjunto é a transposta da representação matricial do operador e usando a simetria do determinante em relação aos seus coeficientes, deduzimos que o determinante do operador transposto é igual ao determinante do operador em si.

TEOREMA 3.8. *Seja  $X$  um espaço vetorial de dimensão finita e considere um operador linear  $\mathbb{T} \in \mathcal{L}(X)$ . Então  $\det(\mathbb{T}) = \det(\mathbb{T}^*)$ .*

DEMONSTRAÇÃO. Seja  $\mathfrak{b} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  uma base de  $X$  e  $\mathfrak{b}' = \{\mathbf{f}_1, \dots, \mathbf{f}_n\}$  a base dual. Segue do [Definição 3.5](#) que

$$[\mathbb{T}]_{\mathfrak{b}} = (a_{ij})_{i,j=1}^n, \quad a_{ij} = \langle \mathbf{f}_i, \mathbb{T}\mathbf{w}_j \rangle.$$

Por outro lado, segue do [Teorema 4.1](#) que a representação  $[\mathbb{T}^*]_{\mathfrak{b}'}$  do operador adjunto na base dual é a transposta de  $[\mathbb{T}]_{\mathfrak{b}}$ , i.e.

$$[\mathbb{T}^*]_{\mathfrak{b}'} = (a_{ji})_{i,j=1}^n = [\mathbb{T}]_{\mathfrak{b}}^{\text{tr}}.$$

Por outro lado, usando a fórmula de Leibniz ([3.1](#)), temos

$$\det(\mathbb{T}) = \sum_{\sigma \in S_n} (-1)^{N(\sigma)} a_{\sigma(1),1} a_{\sigma(2),2} \cdots a_{\sigma(n),n}.$$

Cada permutação  $\sigma \in S_n$  corresponde a uma única permutação inversa  $\sigma^{-1} \in S_n$ , de modo que podemos escrever

$$\det(\mathbb{T}) = \sum_{\sigma \in S_n} (-1)^{N(\sigma)} a_{\sigma(1),\sigma^{-1}(\sigma(1))} a_{\sigma(2),\sigma^{-1}(\sigma(2))} \cdots a_{\sigma(n),\sigma^{-1}(\sigma(n))}.$$

Reordenando  $\sigma(1), \dots, \sigma(n)$  em cada termo, podemos escrever

$$a_{\sigma(1), \sigma^{-1}(\sigma(1))} a_{\sigma(2), \sigma^{-1}(\sigma(2))} \cdots a_{\sigma(n), \sigma^{-1}(\sigma(n))} = a_{1, \sigma^{-1}(1)} a_{2, \sigma^{-1}(2)} \cdots a_{n, \sigma^{-1}(n)}.$$

(Por exemplo,  $a_{21} a_{32} a_{13} = a_{13} a_{32} a_{21}$ , com  $\sigma(1) = 2$ ,  $\sigma(2) = 3$  e  $\sigma(3) = 1$  e  $\sigma^{-1}(1) = 3$ ,  $\sigma^{-1}(2) = 3$  e  $\sigma^{-1}(3) = 2$ ).

Agora, reparametrizando o somatório por  $\xi = \sigma^{-1} \in S_n$ , obtemos

$$\det(\mathbf{T}) = \sum_{\xi \in S_n} a_{1, \xi(1)} a_{2, \xi(2)} \cdots a_{n, \xi(n)} = \det([\mathbf{T}^*]_{\theta'}) = \det(\mathbf{T}^*),$$

concluindo a demonstração. □

## 4. O operador projeção

Um operador especial, chamado de operador projeção, é fundamental no entendimento de quase todos os operadores, sendo útil na “decomposição” de um operador qualquer em partes mais simples. Vamos estudá-lo nesta seção, junto com o operador involutivo, para enriquecer a discussão.

### 4.1. Definição e exemplos.

**DEFINIÇÃO 4.1.** *Seja  $X$  um espaço vetorial. Um operador  $\mathbf{P} \in \mathcal{L}(X)$  é dito uma **projeção** quando  $\mathbf{P}$  é **idempotente**, ou seja,  $\mathbf{P}^2 = \mathbf{P}$ .*

**EXEMPLO 4.1.** Por exemplo, em  $X = \mathbb{R}^2$ , o operador  $\mathbf{P}(x, y) = (x, 0)$  satisfaz  $\mathbf{P}(\mathbf{P}(x, y)) = \mathbf{P}(x, 0) = (x, 0) = \mathbf{P}(x, y)$ , i.e.  $\mathbf{P}^2 = \mathbf{P}$ . O operador  $\mathbf{P}(x, y) = (x - y, 0)$  também satisfaz  $\mathbf{P}^2 = \mathbf{P}$ . Ambos projetam no eixo  $x$ , mas têm núcleos distintos.

**EXEMPLO 4.2.** Observe que o operador nulo  $\mathbf{0}$  e o operador identidade  $\mathbf{I}$  são, também, projeções, ambas triviais, em um certo sentido.

**OBSERVAÇÃO 4.1.** Como será visto na [Seção 4.3](#) e, em particular, na [Definição 4.2](#), dizemos que  $\mathbf{P}$  é um operador projeção sobre  $V = \text{Im}(\mathbf{P})$  ou, mais precisamente, que é o operador projeção sobre  $V = \text{Im}(\mathbf{P})$  ao longo de  $W = \text{ker}(\mathbf{P})$ . Usamos os artigos “um”, no primeiro caso, e “o”, no segundo caso porque pode haver vários operadores projeção sobre um mesmo subespaço  $V$ , com núcleos diferentes, mas só há um operador projeção se especificarmos a imagem e o núcleo.

### 4.2. Propriedades do operador projeção.

**TEOREMA 4.1.** *Dado um operador projeção  $\mathbf{P}$  em um espaço vetorial  $X$ , temos que o espaço se decompõe em soma direta entre o núcleo e a imagem de  $\mathbf{P}$ , i.e.*

$$X = \text{ker}(\mathbf{P}) \oplus \text{Im}(\mathbf{P}).$$

Além disso, todo  $\mathbf{u} \in X$  se escreve de forma única como  $\mathbf{u} = \mathbf{w} + \mathbf{v}$ , onde  $\mathbf{w} \in \ker(\mathbf{P})$  e  $\mathbf{v} \in \text{Im}(\mathbf{P})$  e, com essa decomposição, o operador projeção toma a forma

$$\mathbf{P}(\mathbf{u}) = \mathbf{v}.$$

DEMONSTRAÇÃO. Para ver que é soma direta, devemos mostrar, primeiro, que a interseção é trivial. Suponha, então, que  $\mathbf{u} \in \ker(\mathbf{P}) \cap \text{Im}(\mathbf{P})$ . Como  $\mathbf{u}$  está na imagem de  $\mathbf{P}$ , existe  $\mathbf{v} \in X$  tal que  $\mathbf{u} = \mathbf{P}\mathbf{v}$ . Aplicando  $\mathbf{P}$  a essa identidade e usando que  $\mathbf{P}^2 = \mathbf{P}$ , obtemos que

$$\mathbf{P}\mathbf{u} = \mathbf{P}^2\mathbf{v} = \mathbf{P}\mathbf{v} = \mathbf{u}.$$

Por outro lado,  $\mathbf{u}$  também está no núcleo de  $\mathbf{P}$ , de modo que  $\mathbf{P}\mathbf{u} = \mathbf{0}$ , nos levando a

$$\mathbf{u} = \mathbf{P}\mathbf{u} = \mathbf{0}.$$

Portanto,  $\ker(\mathbf{P}) \cap \text{Im}(\mathbf{P}) = \{\mathbf{0}\}$  é trivial. Isso diz que  $\ker(\mathbf{P}) \oplus \text{Im}(\mathbf{P})$  é soma direta.

Falta deduzir que essa soma direta é o espaço todo. Para isso, usamos o [Teorema 2.1](#) do Núcleo e da Imagem, que nos garante que

$$\dim(X) = \dim(\ker(\mathbf{P})) + \dim(\text{Im}(\mathbf{P})).$$

Como temos uma soma direta entre os espaços acima, vale

$$\begin{aligned} \dim(\ker(\mathbf{P}) + \text{Im}(\mathbf{P})) &= \dim(\ker(\mathbf{P})) + \dim(\text{Im}(\mathbf{P})) - \dim(\ker(\mathbf{P}) \cap \text{Im}(\mathbf{P})) \\ &= \dim(\ker(\mathbf{P})) + \dim(\text{Im}(\mathbf{P})) = \dim(X). \end{aligned}$$

Dessa forma, o subespaço  $\ker(\mathbf{P}) + \text{Im}(\mathbf{P})$  gera o espaço todo  $X$  e obtemos

$$X = \ker(\mathbf{P}) \oplus \text{Im}(\mathbf{P}).$$

Agora, se  $\mathbf{u} = \mathbf{w} + \mathbf{v}$  com  $\mathbf{w} \in \ker(\mathbf{P})$  e  $\mathbf{v} \in \text{Im}(\mathbf{P})$ , então

$$\mathbf{P}\mathbf{u} = \mathbf{P}\mathbf{w} + \mathbf{P}\mathbf{v} = \mathbf{P}\mathbf{v},$$

visto que  $\mathbf{w}$  está no núcleo de  $\mathbf{P}$ . Por sua vez,  $\mathbf{v} = \mathbf{P}\tilde{\mathbf{v}}$ , para algum  $\tilde{\mathbf{v}} \in X$ , visto que  $\mathbf{v}$  está na imagem de  $\mathbf{P}$ . Assim, usando que  $\mathbf{P}$  é idempotente,

$$\mathbf{P}\mathbf{u} = \mathbf{P}^2\tilde{\mathbf{v}} = \mathbf{P}\tilde{\mathbf{v}} = \mathbf{v},$$

completando a demonstração. □

OBSERVAÇÃO 4.2. Na demonstração da [Teorema 4.1](#), é sempre verdade que

$$\dim(\ker(\mathbf{P})) + \dim(\text{Im}(\mathbf{P})) = \dim(X),$$

graças ao Teorema do Núcleo e da Imagem. O que não é sempre verdade é que a interseção  $\ker(\mathbf{P}) \cap \text{Im}(\mathbf{P})$  é trivial. Vejamos, por exemplo,  $\mathbf{T}(x, y) = (y, 0)$ , em  $\mathbb{R}^2$ . Temos  $\ker(\mathbf{T}) = \text{Im}(\mathbf{T}) = \{(x, 0); x \in \mathbb{R}\}$ , que é o eixo  $x$ . Outro exemplo é

$S(x, y) = (y - x, y - x)$ , cuja imagem e o núcleo são dados pela reta  $y = x$ . Na verdade, esses dois operadores são similares entre si.

OBSERVAÇÃO 4.3. O conceito de *idempotente* é mais geral e diz respeito a qualquer elemento  $e$ , em um conjunto  $E$  com uma operação binária  $\circ : E \times E \rightarrow E$ , tal que  $e \circ e = e$ . No caso de operadores lineares, um operador idempotente é uma projeção.

**4.3. Caracterização de um operador projeção.** A soma direta entre o núcleo e a imagem do operador não é o que caracteriza uma projeção, pois vários outros operadores possuem essa propriedade, conforme exemplificado na [Observação 4.4](#).

OBSERVAÇÃO 4.4. Não é necessário ser uma projeção para valer  $\ker(T) \cap \text{Im}(T) = \{\mathbf{0}\}$ . A própria identidade  $I$  é tal que  $\ker(I) = \{\mathbf{0}\}$  e  $\text{Im}(I) = X$ . Qualquer operador invertível também se decompõe assim. Um outro exemplo menos trivial em  $\mathbb{R}^2$  é  $T(x, y) = (2x, 0)$ . A imagem é o eixo  $x$  e o núcleo é o eixo  $y$ , mas temos  $T^2 = 2T$ . E tantos outros exemplos.

No entanto, a decomposição do espaço no [Teorema 4.1](#) junto com a forma  $P\mathbf{u} = \mathbf{v}$  é, sim, suficiente para caracterizar um operador projeção, conforme resultado a seguir.

TEOREMA 4.2. *Seja  $X$  um espaço vetorial e suponha que  $V$  e  $W$  sejam subespaços de  $X$  com*

$$X = W \oplus V.$$

*Então todo  $\mathbf{u} \in X$  se escreve de forma única como  $\mathbf{u} = \mathbf{w} + \mathbf{v}$ , onde  $\mathbf{w} \in W$  e  $\mathbf{v} \in V$ . Isso nos permite definir o operador*

$$P\mathbf{u} = \mathbf{v},$$

*de acordo com essa decomposição. Segue que o operador  $P$  assim definido é uma projeção, com  $\text{Im}(P) = V$  e  $\ker(P) = W$ .*

DEMONSTRAÇÃO. Seja  $\mathbf{u} = \mathbf{w} + \mathbf{v}$ , com  $\mathbf{w} \in W$  e  $\mathbf{v} \in V$ . Pensando na decomposição do próprio  $\mathbf{v}$  entre esses subespaços, vemos que, como ele já está em  $V$ , ele é da forma

$$\mathbf{v} = \mathbf{0} + \mathbf{v},$$

de modo que, pela definição de  $P$ , vale

$$P\mathbf{v} = \mathbf{v}.$$

Sendo assim,

$$P^2\mathbf{u} = P(P\mathbf{u}) = P\mathbf{v} = \mathbf{v} = P\mathbf{u}.$$

Como  $\mathbf{u} \in X$  é arbitrário, provamos que  $P^2 = P$ . □

DEFINIÇÃO 4.2. No contexto do **Teorema 4.2**, dizemos que  $P$  é o **operador projeção sobre  $V$ , ao longo de  $W$** . Se não especificarmos  $W$ , dizemos apenas que  $P$  é um **operador projeção sobre  $V$** .

OBSERVAÇÃO 4.5. No **Teorema 4.2**, o operador  $Q = I - P$  dado por  $Qu = w = u - v$  também é uma projeção. Nesse caso,  $Q$  é o operador projeção em  $W$  ao longo de  $V$ .

EXEMPLO 4.3. Por exemplo, em  $X = \mathbb{R}^2$ , o operador  $P(x, y) = (x, 0)$  é a projeção de  $\mathbb{R}^2$  no eixo  $x$ , ao longo do eixo  $y$ . O seu projetor complementar é  $Q = I - P$ , que se escreve  $Q(x, y) = (x, y) - (x, 0) = (0, y)$ , que é a projeção de  $\mathbb{R}^2$  no eixo  $y$ , ao longo do eixo  $x$ .

EXEMPLO 4.4. O operador  $P(x, y) = (x - y, 0)$  também projeta no eixo  $x$ , mas ao longo da reta  $y = x$ . O seu projetor complementar  $Q(x, y) = I(x, y) - P(x, y) = (x, y) - (x - y, 0) = (y, y)$  projeta na reta  $y = x$ , ao longo do eixo  $x$ . A representação de  $P$  na base canônica  $e = \{(1, 0), (0, 1)\}$  é

$$[P]_e = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix},$$

enquanto que

$$[Q]_e = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}.$$

#### 4.4. Representação matricial de um operador projeção.

TEOREMA 4.3. Dado um operador projeção  $P$  em um espaço vetorial  $X$  de dimensão finita  $n = \dim(X) \in \mathbb{N}$ , então, em uma base  $\mathcal{b}$  apropriada, temos a representação

$$[P]_{\mathcal{b}} = \begin{bmatrix} \mathbf{I}_m & \mathbf{O}_{m, n-m} \\ \mathbf{O}_{n-m, m} & \mathbf{O}_{n-m, n-m} \end{bmatrix},$$

onde  $m = \dim(\text{Im}(T))$  é o posto de  $T$ ,  $0 \leq m \leq n$ .

**4.5. Operadores involutivos.** Um tipo simples de operador que pode ser entendido facilmente junto com o operador projeção é o de operadores involutivos.

DEFINIÇÃO 4.3. Seja  $X$  um espaço vetorial. Um operador  $R \in \mathcal{L}(X)$  é dito **involutivo** quando satisfaz a identidade  $R^2 = I$ . Em outras palavras, quando  $R$  é invertível é a sua própria inversa,  $R^{-1} = R$ .

EXEMPLO 4.5 (Reflexões). Por exemplo, em  $X = \mathbb{R}^2$ , i.e. no plano  $xy$ , a reflexão em relação ao eixo  $x$ , dada por  $R(x, y) = (x, -y)$  é um operador involutivo. Assim como a reflexão  $R(x, y) = (-x, -y)$  em relação à origem. Mas nem todo operador involutivo é uma reflexão. A reflexão envolve a ideia de ortogonalidade, de modo

que a definição do que é um operador reflexão, no sentido mais geral, será vista posteriormente.

EXEMPLO 4.6. Ainda em  $X = \mathbb{R}^2$ , o operador  $R(x, y) = (x - 2y, -y)$  satisfaz  $R(R(x, y)) = R(x - 2y, -y) = ((x - 2y) - 2(-y), -(-y)) = (x, y) = I(x, y)$ , ou seja, também é involutivo. Se assemelha a uma reflexão, “rebatendo” no eixo  $x$ , mas o faz de forma paralela à reta  $y = x$ , que não é ortogonal ao eixo  $x$ . Em particular, a norma Euclidiana do vetor  $R(x, y)$ , proveniente do produto interno, é, em geral, diferente da norma de  $(x, y)$ . Novamente, só trabalharemos com produto interno mais pra frente, mas aproveitamos para já colocar essa diferença em perspectiva.

TEOREMA 4.4. *Dado um operador involutivo  $R$  em um espaço vetorial  $X$ , temos que o espaço se decompõe em soma direta entre os vetores mantidos fixos e os vetores levados em seus vetores inversos, i.e.*

$$X = V \oplus W,$$

onde

$$V = \{\mathbf{u} \in X; R\mathbf{u} = \mathbf{u}\}, \quad W = \{\mathbf{u} \in X; R\mathbf{u} = -\mathbf{u}\}$$

Além disso, todo  $\mathbf{u} \in X$  se escreve de forma única como  $\mathbf{u} = \mathbf{v} + \mathbf{w}$ , onde  $\mathbf{v} \in V$  e  $\mathbf{w} \in W$  e, assim, o operador  $R$  age da forma

$$R(\mathbf{u}) = R(\mathbf{v} + \mathbf{w}) = \mathbf{v} - \mathbf{w}.$$

DEMONSTRAÇÃO. Em primeiro lugar, é imediato verificar que  $V$  e  $W$  são subespaços vetoriais. Agora, se  $\mathbf{u} \in V \cap W$ , então, por estar em  $V$ , vale  $R\mathbf{u} = \mathbf{u}$ . Por outro lado, como também está em  $W$ , então  $R\mathbf{u} = -\mathbf{u}$ . Logo,  $\mathbf{u} = R\mathbf{u} = -\mathbf{u}$ , de modo que  $\mathbf{u} = \mathbf{0}$ . Assim,  $V \cap W = \{\mathbf{0}\}$  e temos uma soma direta  $V \oplus W$ . Para ver que  $V + W$  gera todo o  $X$ , seja  $\mathbf{u} \in X$ . Somando e subtraindo  $(1/2)R\mathbf{u}$ , temos

$$\mathbf{u} = \mathbf{u} + \frac{1}{2}R\mathbf{u} - \frac{1}{2}R\mathbf{u} = \frac{\mathbf{u} + R\mathbf{u}}{2} + \frac{\mathbf{u} - R\mathbf{u}}{2}.$$

Sejam

$$\mathbf{v} = \frac{\mathbf{u} + R\mathbf{u}}{2}, \quad \mathbf{w} = \frac{\mathbf{u} - R\mathbf{u}}{2}.$$

Vamos verificar que  $\mathbf{v} \in V$  e  $\mathbf{w} \in W$ . Para isso, observe que, usando que  $R^2 = I$ ,

$$R\mathbf{v} = \frac{R\mathbf{u} + R^2\mathbf{u}}{2} = \frac{R\mathbf{u} + \mathbf{u}}{2} = \mathbf{v},$$

de modo que  $\mathbf{v} \in V$ . Da mesma forma,

$$R\mathbf{w} = \frac{R\mathbf{u} - R^2\mathbf{u}}{2} = \frac{R\mathbf{u} - \mathbf{u}}{2} = -\frac{\mathbf{u} - R\mathbf{u}}{2} = -\mathbf{w},$$

de modo que  $\mathbf{w} \in W$ . Com isso, provamos a soma direta

$$X = V \oplus W.$$

Agora, se  $\mathbf{u} = \mathbf{v} + \mathbf{w}$  com  $\mathbf{v} \in V$  e  $\mathbf{w} \in W$ , então, usando a linearidade e a propriedade desses subespaços, obtemos que

$$R\mathbf{u} = R(\mathbf{v} + \mathbf{w}) = R\mathbf{v} + R\mathbf{w} = \mathbf{v} - \mathbf{w}.$$

Isso completa a demonstração.  $\square$

De maneira análoga ao do operador projeção, temos o seguinte resultado.

**TEOREMA 4.5.** *Seja  $X$  um espaço vetorial e suponha que  $V$  e  $W$  sejam subespaços de  $X$  com*

$$X = V \oplus W.$$

*Então todo  $\mathbf{u} \in X$  se escreve de forma única como  $\mathbf{u} = \mathbf{w} + \mathbf{v}$ , onde  $\mathbf{w} \in W$  e  $\mathbf{v} \in V$ . Isso nos permite definir o operador linear  $R : X \rightarrow X$  dado por*

$$R\mathbf{u} = R(\mathbf{v} + \mathbf{w}) = \mathbf{v} - \mathbf{w},$$

*de acordo com essa decomposição. Segue que o operador  $R$  assim definido é involutivo.*

**DEMONSTRAÇÃO.** Seja  $\mathbf{u} = \mathbf{v} + \mathbf{w}$ , com  $\mathbf{v} \in W$  e  $\mathbf{w} \in V$ . Então, pela definição de  $R$ , temos

$$R\mathbf{u} = \mathbf{v} - \mathbf{w}.$$

Como  $\mathbf{v} \in V$  e  $-\mathbf{w} \in W$  e a decomposição é única, então aplicando novamente  $R$ , obtemos

$$R^2\mathbf{u} = R(\mathbf{v} - \mathbf{w}) = \mathbf{v} - (-\mathbf{w}) = \mathbf{v} + \mathbf{w} = \mathbf{u}.$$

Como  $\mathbf{u} \in X$  é arbitrário, provamos que  $R^2 = I$ , ou seja, que  $R$  é involutivo.  $\square$

## 5. Subespaços invariantes

O conceito de subespaço é natural no estudo de espaços vetoriais. Quanto combinamos isso com operadores lineares, é natural questionar como um operador linear opera em cada subespaço do espaço vetorial. É razoável esperar, também, que certos subespaços tenham propriedades especiais em relação à ação do operador linear. Em particular, podemos ter um subespaço sendo levado nele próprio. Isso pode ser explorado no sentido de reduzir o entendimento do operador ao de sua restrição a esses subespaços. Esses espaços, ditos invariantes, têm um papel fundamental no entendimento de operadores lineares e, também, por outro lado, na construção de operadores lineares com certas propriedades desejadas. Esses espaços tem uma conexão direta com os operadores de projeção, como veremos a seguir.

### 5.1. Definição.

DEFINIÇÃO 5.1. *Seja  $T \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$ . Um **subespaço invariante** de  $T$  é um subespaço  $U \subset X$  tal que*

$$T(U) \subset U.$$

Restrito a  $U$ , o operador mantém, obviamente, a sua propriedade de linearidade. Dessa forma, podemos definir um “novo” operador linear apenas restringindo o operador ao subespaço invariante.

DEFINIÇÃO 5.2. *Seja  $T \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  e suponha que  $U \subset X$  seja um subespaço invariante de  $T$ . Então podemos definir a **restrição de  $T$  a  $U$**  como o operador linear*

$$T|_U : U \rightarrow U$$

dado por

$$T|_U(\mathbf{u}) = T(\mathbf{u}), \quad \forall \mathbf{u} \in U.$$

É claro que o próprio  $X$  e o subespaço trivial  $\{\mathbf{0}\}$  podem ser vistos como subespaços invariantes, mas o interesse maior está, obviamente, nos subespaços próprios não triviais. Vejamos alguns exemplos.

EXEMPLO 5.1. O operador  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  dado por  $T(x, y) = (2x, 3y)$ , para  $(x, y) \in \mathbb{R}^2$ , possui dois espaços vetoriais próprios não triviais, o eixo  $x$  e o eixo  $y$ . Ao longo do eixo  $x$ , o operador dobra o tamanho dos vetores,  $T|_{\text{eixo } x} \mathbf{u} = 2\mathbf{u}$ , enquanto que no eixo  $y$ , o operador triplica os vetores de tamanho,  $T|_{\text{eixo } y} \mathbf{u} = 3\mathbf{u}$ .

EXEMPLO 5.2. No caso do operador identidade, qualquer subespaço vetorial é invariante.

EXEMPLO 5.3. Para qualquer operador linear  $T \in \mathcal{L}(X)$ , em um espaço vetorial arbitário  $X$ , não é difícil verificar que tanto o núcleo como a imagem de  $T$  são invariantes. Naturalmente, quando restrito ao núcleo, o operador se reduz ao operador nulo, i.e.  $T|_{\ker(T)} = \mathbf{0}$  (onde no lado direito temos o operador nulo no espaço  $\ker(T)$ ), que, por sua vez, também pode ser visto como a restrição do operador nulo de  $X$  ao núcleo de  $T$ . Ou seja, podemos escrever, também,  $T|_{\ker(T)} = \mathbf{0}|_{\ker(T)}$ , onde agora  $\mathbf{0}$  se refere ao operador nulo em  $X$ .

EXEMPLO 5.4. O operador  $T(x, y) = (x+y, x+y)$  tem núcleo  $\{y = -x\}$  e imagem  $\{y = x\}$ . Como todo subespaço invariante deve pertencer à imagem do operador, o único subespaço não trivial capaz de ser invariante é a reta  $y = x$ . E, de fato, se  $y = x$ , então  $T(x, y) = T(x, x) = (2x, 2x)$ , que também está nessa reta.  $T|_{y=x} \mathbf{u} = 2\mathbf{u}$

EXEMPLO 5.5. No caso de um operador projeção  $P : X \rightarrow Y$ , visto que  $P^2 = P$ , temos  $P\mathbf{u} = \mathbf{u}$ , para todo  $\mathbf{u} \in \text{Im}(P)$ , ou seja, a restrição de  $P$  à sua imagem é a identidade. Podemos escrever isso como  $P|_{\text{Im}(P)} = I_{\text{Im}(P)}$ , onde o subscrito de  $I$  indica em qual espaço o operador identidade age. Por outro lado, assim como para qualquer outro operador,  $P|_{\text{ker}(P)} = 0_{\text{ker}(P)}$ .

EXEMPLO 5.6. No espaço dos polinômios  $\mathcal{P}$ , os subespaços  $\mathcal{P}_n$  dos polinômios com grau no máximo  $n \in \mathbb{N}$  são invariantes pelo operador derivada  $p(x) \mapsto p'(x)$ . Mas esses subespaços não são invariantes pelo operador integração  $p(x) \mapsto \int_0^x p(s) ds$ .

**5.2. Conexão com o operador projeção.** O próximo resultado diz que se um operador linear  $T$  comuta com um determinado operador projeção, então o subespaço sobre o qual o espaço é projetado e o subespaço ao longo do qual o espaço é projetado são invariantes pelo operador  $T$ . Isso tem um papel crucial na teoria espectral e no entendimento de operadores lineares.

TEOREMA 5.1. *Seja  $T \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  e seja  $P \in \mathcal{L}(X)$  um operador projeção. Se  $TP = PT$ , então  $V = \text{Im}(P)$  e  $W = \text{ker}(P)$  são subespaços invariantes de  $T$ .*

DEMONSTRAÇÃO. Se  $\mathbf{u} \in V = \text{Im}(P)$ , temos  $\mathbf{u} = P\mathbf{u}$ . Usando a comutatividade, deduzimos que

$$T\mathbf{u} = TP\mathbf{u} = PT\mathbf{u} \in \text{Im}(P) = V,$$

ou seja,  $TV \subset V$ , provando que  $V$  é invariante por  $T$ .

Agora, seja  $Q = I - P$ . Então,  $Q$  também comuta com  $T$ , visto que

$$TQ = T(I - P) = T - TP = T - PT = (I - P)T = QT.$$

Sendo  $Q$  um operador projeção que comuta com  $T$ , segue da primeira parte da demonstração que  $\text{Im}(Q)$  é invariante por  $T$ . Mas  $\text{Im}(Q) = \text{ker}(P)$ , de modo que  $W = \text{ker}(P)$  também é invariante.  $\square$

OBSERVAÇÃO 5.1. A comutatividade com o operador projeção, trabalhada no Teorema 5.1 não é, no entanto, estritamente necessária para um subespaço ser invariante. Podemos ter diferentes projeções sobre um mesmo subespaço invariante que não comutem com o operador. Mais ainda, pode nem haver uma projeção que comute com o operador, conforme veremos nos exemplos a seguir. Mas se consideramos um subespaço invariante em conjunto com um subespaço complementar também invariante, então a comutatividade se torna uma caracterização. Isso é visto no Exercício 7.11. Ou seja, assumindo que  $X = S \oplus S'$  e que  $P$  é a projeção sobre  $S$  ao longo de  $S'$ , então  $S$  e  $S'$  são invariantes se, e somente se,  $T$  e  $P$  comutam. Um resultado análogo vale para  $X = S_1 \oplus \cdots \oplus S_m$ . Veja Exercício 7.12

EXEMPLO 5.7. Considere o operador linear  $T(x, y) = (2x, 3y)$ , visto no **Exemplo 5.1**. Os eixos  $x$  e  $y$  são invariantes. Se escolhermos a projeção  $P(x, y) = (x, 0)$  sobre o eixo  $x$ , ao longo do eixo  $y$  (ambos invariantes), cuja projeção complementar é  $Q(x, y) = (0, y)$ , então é fácil ver que

$$\begin{aligned} (TP)(x, y) &= T(x, 0) = (2x, 0), \\ (PT)(x, y) &= P(2x, 3y) = (2x, 0), \\ (TQ)(x, y) &= T(0, y) = (0, 3y), \\ (QT)(x, y) &= Q(2x, 3y) = (0, 3y), \end{aligned}$$

ou seja,  $TP = PT$  e  $TQ = QT$ .

Por outro lado, se escolhermos a projeção sobre o eixo  $x$  (invariante) ao longo da reta  $y = x$  (que não é invariante), i.e.  $P(x, y) = (x - y, 0)$ , com  $Q(x, y) = (x, y) - (x - y, 0) = (y, y)$ , então

$$\begin{aligned} (TP)(x, y) &= T(x - y, 0) = (2(x - y), 0), \\ (PT)(x, y) &= P(2x, 3y) = (2x - 3y, 0), \\ (TQ)(x, y) &= T(y, y) = (2y, 3y), \\ (QT)(x, y) &= Q(2x, 3y) = (3y, 3y), \end{aligned}$$

ou seja  $TP \neq PT$  e  $TQ \neq QT$ .

EXEMPLO 5.8. Em alguns casos, pode nem ser questão de escolha. Pode simplesmente não haver projeção que comuta com o operador. Por exemplo, considerando  $T(x, y) = (y, 0)$ , então qualquer projeção  $P$  sobre o eixo  $x$  é tal que

$$TP(x, y) = T(z, 0) = (0, 0),$$

para algum  $z$  intermediário, enquanto que

$$PT(x, y) = P(y, 0) = (y, 0),$$

visto que  $P$  é a identidade ao longo do eixo  $x$ .

EXEMPLO 5.9. No exemplo  $T(x, y) = (2x, 3y)$ , considerado no **Exemplo 5.1** e no **Exemplo 5.7**, definindo  $P(x, y) = (x, 0)$ , com complementar  $Q(x, y) = (0, y)$ , observe que podemos escrever

$$T = 2P + 3Q,$$

visto que

$$(2P + 3Q)(x, y) = 2P(x, y) + 3Q(x, y) = 2(x, 0) + 3(0, y) = (2x, 3y) = T(x, y).$$

Esso tipo de decomposição será buscada plenamente na parte de teoria espectral.

**5.3. Representação matricial em blocos.** Suponha que  $X = U \oplus V$  e sejam  $P$  e  $Q$  as projeções sobre  $U$  e  $V$ , ao longo de  $V$  e  $U$ , respectivamente. Dada  $T \in \mathcal{L}(X)$ , como  $I = P + Q$ , podemos escrever

$$T = (P + Q)T(P + Q) = PTP + PTQ + QTP + QTQ.$$

Aplicando  $P$  e usando que  $P^2 = I$  e  $PQ = 0$ , temos

$$PT = PTP + PTQ.$$

Aplicando agora  $Q$ , obtemos, analogamente,

$$QT = QTP + QTQ.$$

Em forma de sistema, temos

$$\begin{cases} PT = PTP + PTQ, \\ QT = QTP + QTQ. \end{cases}$$

Essas equações de operadores podem ser escritas na “forma matricial”

$$\begin{pmatrix} PT \\ QT \end{pmatrix} = \begin{bmatrix} PTP & PTQ \\ QTP & QTQ \end{bmatrix} \begin{pmatrix} P \\ Q \end{pmatrix}.$$

As projeções  $P$  e  $Q$  que aparecem à direita nos coeficientes da matriz acima parecem redundantes, já que serão multiplicadas por  $P$  e  $Q$ , respectivamente, mas a inclusão delas deixa mais claro o operador que age em cada bloco e tem uma ligação mais direta com a representação final do operador em blocos, uma vez escolhida uma base de  $X$  a partir de bases de  $U$  e  $V$ .

De fato, dada uma base  $\mathcal{b}_1$  de  $U$  e uma base  $\mathcal{b}_2$  de  $V$ , então  $\mathcal{b} = \mathcal{b}_1 \cup \mathcal{b}_2$  é base de  $X$  e temos, em nível matricial,

$$[T]_{\mathcal{b}} = \begin{pmatrix} [PT|_U]_{\mathcal{b}_1} \\ [QT|_V]_{\mathcal{b}_2} \end{pmatrix} = \begin{pmatrix} [PT|_U]_{\mathcal{b}_1}^{\mathcal{b}_1} + [QT|_V]_{\mathcal{b}_1}^{\mathcal{b}_2} \\ [QT|_V]_{\mathcal{b}_2}^{\mathcal{b}_1} + [QT|_V]_{\mathcal{b}_2}^{\mathcal{b}_2} \end{pmatrix},$$

enquanto que, em nível vetorial,

$$\begin{aligned} [T\mathbf{u}]_{\mathcal{b}} &= [T]_{\mathcal{b}}[\mathbf{u}]_{\mathcal{b}} = [T]_{\mathcal{b}} \begin{pmatrix} [P\mathbf{u}]_{\mathcal{b}_1} \\ [Q\mathbf{u}]_{\mathcal{b}_2} \end{pmatrix} = \begin{pmatrix} [PT|_U]_{\mathcal{b}_1}^{\mathcal{b}_1} [P\mathbf{u}]_{\mathcal{b}_1} \\ [QT|_V]_{\mathcal{b}_2}^{\mathcal{b}_1} [P\mathbf{u}]_{\mathcal{b}_1} + [QT|_V]_{\mathcal{b}_2}^{\mathcal{b}_2} [Q\mathbf{u}]_{\mathcal{b}_2} \end{pmatrix} \\ &= \begin{bmatrix} [PT|_U]_{\mathcal{b}_1}^{\mathcal{b}_1} & [PT|_V]_{\mathcal{b}_1}^{\mathcal{b}_2} \\ [QT|_U]_{\mathcal{b}_2}^{\mathcal{b}_1} & [QT|_V]_{\mathcal{b}_2}^{\mathcal{b}_2} \end{bmatrix} \begin{pmatrix} [P\mathbf{u}]_{\mathcal{b}_1} \\ [Q\mathbf{u}]_{\mathcal{b}_2} \end{pmatrix}. \end{aligned}$$

Voltando para a representação

$$\begin{pmatrix} PT \\ QT \end{pmatrix} = \begin{bmatrix} PTP & PTQ \\ QTP & QTQ \end{bmatrix} \begin{pmatrix} P \\ Q \end{pmatrix}$$

estas não são, exatamente, matrizes reais ou complexas, mas vemos que a mesma estrutura de blocos e multiplicação em blocos vista no [Exemplo 1.20](#) vale.

Agora, se  $U = PX$  é invariante, então  $TPX \in U$ , de modo que  $QTP = 0$ . Nesse caso, temos a forma triangular

$$\begin{cases} PT = PTP + PTQ, \\ QT = \quad \quad QTQ, \end{cases}$$

que pode ser escrita como

$$\begin{pmatrix} PT \\ QT \end{pmatrix} = \begin{bmatrix} PTP & PTQ \\ 0 & QTQ \end{bmatrix} \begin{pmatrix} P \\ Q \end{pmatrix}$$

Se, além disso,  $V = QX$  também é invariante, então temos a forma diagonal

$$\begin{cases} PT = PTP \\ QT = \quad \quad QTQ. \end{cases}$$

que pode ser escrita como

$$\begin{pmatrix} PT \\ QT \end{pmatrix} = \begin{bmatrix} PTP & 0 \\ 0 & QTQ \end{bmatrix} \begin{pmatrix} P \\ Q \end{pmatrix}$$

Para fazer uma conexão mais concreta com matrizes, escrevemos isso da seguinte forma

**TEOREMA 5.2.** *Seja  $T \in \mathcal{L}(X)$  um operador linear em um espaço vetorial não trivial  $X$  de dimensão finita. Suponha que  $X = V \oplus W$ , com subespaços não triviais  $V$  e  $W$ , e sejam  $\mathcal{b}_1 = \{\mathbf{w}_1, \dots, \mathbf{w}_{n_1}\}$  e  $\mathcal{b}_2 = \{\mathbf{w}_{n_1+1}, \dots, \mathbf{w}_{n_1+n_2}\}$  bases de  $V$  e  $W$ , onde  $n_1 = \dim(V)$  e  $n_2 = \dim(W)$ . Considere a base  $\mathcal{b} = \mathcal{b}_1 \cup \mathcal{b}_2$  de  $X$ . Então*

$$[T]_{\mathcal{b}} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}, \quad (5.1)$$

onde

$$\mathbf{A} = [PT|_V]_{\mathcal{b}_1}, \quad \mathbf{B} = [PT|_W]_{\mathcal{b}_1}^{\mathcal{b}_2}, \quad \mathbf{C} = [QT|_V]_{\mathcal{b}_2}^{\mathcal{b}_1}, \quad \mathbf{D} = [QT|_W]_{\mathcal{b}_2}.$$

Se  $V$  for subespaço invariante de  $T$ , então

$$[T]_{\mathcal{b}} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{D} \end{bmatrix}. \quad (5.2)$$

Se  $W$  for subespaço invariante de  $T$ , então

$$[T]_{\mathcal{b}} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}. \quad (5.3)$$

Se ambos  $V$  e  $W$  forem subespaços invariantes de  $T$ , então

$$[T]_{\mathfrak{B}} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix}. \quad (5.4)$$

DEMONSTRAÇÃO. Deixamos a verificação desse resultado para o leitor.  $\square$

OBSERVAÇÃO 5.2. Para simplificar a visualização, blocos nulos podem ser omitidos. Assim, (5.2), (5.3) e (5.4) tomam as formas

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ & \mathbf{D} \end{bmatrix}, \quad \begin{bmatrix} \mathbf{A} & \\ \mathbf{C} & \mathbf{D} \end{bmatrix}, \quad \begin{bmatrix} \mathbf{A} & \\ & \mathbf{D} \end{bmatrix}.$$

EXEMPLO 5.10. Considere o operador  $T : \mathbb{R}^4 \rightarrow \mathbb{R}^4$  dado por  $T(x, y, z, w) = (x + 2y + 3z + 4w, y + 2z + 3w, z + 2w, w)$ . Então o eixo  $x$  é invariante, o plano  $xy$  é invariante e o espaço  $xyz$  é invariante. Escrevendo  $\mathbb{R}^4 = \{\text{plano } xy\} \times \{\text{plano } zw\}$ , temos, usando a base canônica,

$$\left[ \begin{array}{cc|cc} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ \hline 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{array} \right]$$

ou, omitindo os blocos nulos,

$$\left[ \begin{array}{cc|cc} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ \hline & & 1 & 2 \\ & & 0 & 1 \end{array} \right].$$

EXEMPLO 5.11. Considere o operador  $T : \mathbb{R}^4 \rightarrow \mathbb{R}^4$  dado por  $T(x, y, z, w) = (x + y, x + 2y, z + w, z + 3w)$ . Então os subespaços  $xy$  e  $zw$  são invariantes. Na base canônica temos

$$\left[ \begin{array}{cc|cc} 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ \hline 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 3 \end{array} \right]$$

ou, omitindo os blocos nulos,

$$\left[ \begin{array}{cc|cc} 1 & 1 & & \\ 1 & 2 & & \\ \hline & & 1 & 1 \\ & & 0 & 3 \end{array} \right].$$

**5.4. Subespaços encaixantes e forma triangular em blocos.** As representações (5.2) e (5.3) do Teorema 5.2 podem ser estendidas a mais de dois subespaços invariantes. A chave para obter uma forma triangular é que esses subespaços sejam *encaixantes*.

TEOREMA 5.3. *Seja  $T \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  de dimensão finita  $n \in \mathbb{N}$ . Suponha que  $S_1, \dots, S_m$ ,  $m \in \mathbb{N}$ ,  $1 \leq m \leq n$ , sejam subespaços invariantes de  $T$ , com*

$$\{0\} \subsetneq S_1 \subsetneq S_2 \subsetneq \dots \subsetneq S_m = X.$$

*Seja  $\mathcal{b} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  uma base de  $X$  associada a esses subespaços encaixantes, i.e. tal que*

$$S_j = \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_{n_j}\},$$

*onde  $n_j = \dim(S_j)$ ,*

$$1 \leq n_1 < n_2 < \dots < n_m = n.$$

*Então a representação de  $T$  na base  $\mathcal{b}$  tem uma forma triangular superior em blocos, i.e.*

$$[T]_{\mathcal{b}} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1m} \\ & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2m} \\ & & \ddots & \vdots \\ & & & \mathbf{A}_{mm} \end{bmatrix},$$

*onde  $\mathbf{A}_{jj}$  são blocos de matrizes quadradas de dimensões  $(n_j - n_{j-1}) \times (n_j - n_{j-1})$  e  $\mathbf{A}_{ij}$  são blocos de matrizes retangulares de dimensões  $(n_i - n_{i-1}) \times (n_j - n_{j-1})$ .*

DEMONSTRAÇÃO. Considerando a base dual  $\{\mathbf{f}_1, \dots, \mathbf{f}_n\}$ , cada coeficiente de  $[T]_{\mathcal{b}} = (a_{ij})_{i,j=1}^{i,j=n}$  é dado por  $a_{ij} = \langle \mathbf{f}_i, T\mathbf{w}_j \rangle$ . Primeiro, observamos que cada bloco é dado por

$$\mathbf{A}_{ij} = (\langle \mathbf{f}_{i'}, T\mathbf{w}_{j'} \rangle)_{i',j'}, \quad n_{i-1} < i' \leq n_i, \quad n_{j-1} < j' \leq n_j,$$

para  $i, j = 1, \dots, m$ . Em seguida, para cada  $n_{j-1} < j' \leq n_j$ , temos  $\mathbf{w}_{j'} \in S_j \setminus S_{j-1}$  e

$$T\mathbf{w}_{j'} \in \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_{n_j}\},$$

de modo que

$$\langle \mathbf{f}_{i'}, T\mathbf{w}_{j'} \rangle = 0, \quad i' > n_j.$$

Ou seja, os blocos  $\mathbf{A}_{ij}$  com  $i > j$  são todos nulos, restando a forma triangular superior.  $\square$

OBSERVAÇÃO 5.3. Uma forma triangular inferior pode ser obtida a partir das condições do Teorema 5.3, apenas reordenando a base. Tomando, por exemplo,  $\mathcal{a} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  com  $S_j = \text{span}\{\mathbf{v}_{n-n_j}, \dots, \mathbf{v}_n\}$ , obtemos  $[\mathbf{T}]_{\mathcal{a}}$  da forma

$$[\mathbf{T}]_{\mathcal{a}} = \begin{bmatrix} \mathbf{A}_{11} & & & \\ \mathbf{A}_{21} & \mathbf{A}_{22} & & \\ \vdots & \vdots & \ddots & \\ \mathbf{A}_{m1} & \mathbf{A}_{m2} & \cdots & \mathbf{A}_{mm} \end{bmatrix},$$

EXEMPLO 5.12. Considere o operador  $\mathbf{T} : \mathbb{R}^4 \rightarrow \mathbb{R}^4$  dado por

$$\mathbf{T}(x, y, z) = \mathbf{T}_{\mathbf{A}}(x, y, z) = (x - y, x - z, x - w, w - y),$$

associado à matriz

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

Observe que

$$\mathbf{T}(1, 1, 1, 1) = (0, 0, 0, 0) = 0(1, 1, 1, 1).$$

e

$$\mathbf{T}(1, 0, 1, 0) = (1, 0, 1, 0)$$

Para completar a base, escolhemos vetores canônicos,

$$\mathcal{b} = \{(1, 1, 1, 1), (1, 0, 1, 0), (0, 0, 1, 0), (0, 0, 0, 1)\}.$$

Temos

$$\mathbf{T}(0, 0, 1, 0) = (0, -1, 0, 0) = -(1, 1, 1, 1) + (1, 0, 1, 0) + (0, 0, 0, 1)$$

e

$$\mathbf{T}(0, 0, 0, 1) = (0, 0, -1, 1) = -(0, 0, 1, 0) + (0, 0, 0, 1).$$

Assim, obtemos a representação triangular em blocos

$$[\mathbf{T}]_{\mathcal{b}} = \left[ \begin{array}{cc|cc} 0 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ \hline 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 1 \end{array} \right]$$

### 5.5. Subespaços invariantes independentes e forma diagonal em blocos.

A representação (5.4) do Teorema 5.2 também pode ser estendida a mais de dois subespaços invariantes. A chave para obter uma forma diagonal em blocos é que esses subespaços sejam linearmente independentes entre si.

TEOREMA 5.4. *Seja  $T \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  de dimensão finita  $n \in \mathbb{N}$ . Suponha que  $S_1, \dots, S_m$ ,  $m \in \mathbb{N}$ ,  $1 \leq m \leq n$ , sejam subespaços invariantes de  $T$  e linearmente independentes. Seja  $\mathcal{B} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  uma base de  $X$  associada a esses subespaços independentes, i.e. tal que*

$$S_j = \text{span}\{\mathbf{w}_{n_1+\dots+n_{j-1}+1}, \dots, \mathbf{w}_{n_1+\dots+n_{j-1}+n_j}\},$$

onde  $n_j = \dim(S_j)$ ,

$$n_1 + n_2 + \dots + n_m = n.$$

Então a representação de  $T$  na base  $\mathcal{B}$  tem uma forma diagonal em blocos, i.e.

$$[T]_{\mathcal{B}} = \begin{bmatrix} \mathbf{A}_{11} & & & \\ & \mathbf{A}_{22} & & \\ & & \ddots & \\ & & & \mathbf{A}_{mm} \end{bmatrix},$$

onde  $\mathbf{A}_{jj}$  são blocos de matrizes quadradas de dimensões  $n_j \times n_j$ .

TEOREMA 5.5. *No contexto do Teorema 5.4, também podemos escrever*

$$T = TP_1 + \dots + TP_m = P_1T + \dots + P_mT,$$

onde  $P_j$  são as projeções sobre cada  $S_j$ , ao longo de  $\bigoplus_{i \neq j} S_i$ .

DEMONSTRAÇÃO. Deixamos a demonstração para o leitor. Em relação a isso, veja o Exercício 7.12.  $\square$

EXEMPLO 5.13. Considere, novamente, o operador do Exemplo 2.1, dado por

$$T_{\mathbf{A}}(x, y, z) = (2x - y + z, x + z, x - y + 2z),$$

e considere os subespaços invariantes

$$S_1 = \text{span}\{(1, 1, 0)\}, \quad S_2 = \text{span}\{(0, 1, 1)\}, \quad S_3 = \text{span}\{(1, 1, 1)\}.$$

Como podemos escrever um vetor  $\mathbf{u} = (x, y, z)$  na base canônica na base  $\mathcal{B}$ ? Resolvendo

$$(x, y, z) = a(1, 1, 0) + b(0, 1, 1) + c(1, 1, 1) = (a + c, a + b + c, b + c),$$

ou seja, precisamos resolver o sistema

$$\begin{cases} a & + c = x, \\ a + b + c = y, \\ & b + c = z, \end{cases}$$

que nos dá  $a = y - z$ ,  $b = x + z - y$ ,  $c = y - x$ , ou seja

$$(x, y, z) = (y - z)(1, 1, 0) + (y - x)(0, 1, 1) + (x - y + z)(1, 1, 1),$$

ou

$$[(x, y, z)]_{\mathcal{B}} = (y - z, y - x, x + z - y).$$

Observe, agora, que as projeções sobre cada subespaço, ao longo dos outros subespaços, são obtidas retendo apenas a componente em cada subespaço, i.e.

$$\mathbf{P}_1(x, y, z) = (y - z)(1, 1, 0) = (y - z, y - z, 0),$$

$$\mathbf{P}_2(x, y, z) = (y - x)(0, 1, 1) = (0, y - x, y - x),$$

$$\mathbf{P}_3(x, y, z) = (x + z - y)(1, 1, 1) = (x + z - y, x + z - y, x + z - y).$$

Dessa forma, podemos escrever o operador como

$$\mathbf{T} = \mathbf{P}_1 + \mathbf{P}_2 + 2\mathbf{P}_3.$$

Na forma matricial, podemos escrever

$$\begin{aligned} \mathbf{A} &= \mathbf{P}_1 + \mathbf{P}_2 + \mathbf{P}_3 \\ &= \begin{bmatrix} 0 & 1 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 1 & 0 \end{bmatrix} + 2 \begin{bmatrix} 1 & -1 & 1 \\ 1 & -1 & 1 \\ 1 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & 2 \end{bmatrix}. \end{aligned}$$

onde  $\mathbf{A} = [\mathbf{T}]_{\mathcal{E}}$  e  $\mathbf{P}_j = [\mathbf{P}_j]_{\mathcal{E}}$ ,  $j = 1, 2, 3$ .

## 6. Aplicações

**6.1. Transformada discreta de Fourier.** Por vir...

**6.2. Operador de Koopman em sistemas dinâmicos.** Por vir...

## 7. Exercícios

## Exercícios

- 7.1. Seja  $T : X \rightarrow X$  um operador linear em um espaço vetorial  $X$ .
- (1) Mostre que se  $\mathbf{u} \in \text{Im}(T)$ , então  $T\mathbf{u} \in \text{Im}(T)$ .
  - (2) Mostre que se  $\mathbf{u} \in \text{ker}(T)$ , então  $T\mathbf{u} \in \text{ker}(T)$ .
- 7.2. Encontre um operador  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  tal que o núcleo e a imagem de  $T$  sejam iguais à reta  $y = 2x$ .
- 7.3. Encontre a projeção  $P : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  do espaço  $\mathbb{R}^3$ , com coordenadas  $xyz$ , sobre o plano  $xy$ , ao longo da reta  $z = y = x$ . Encontre, também, a projeção complementar  $Q = I - P$ .
- 7.4. Encontre a projeção  $P : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  de  $\mathbb{R}^3$  sobre o subespaço

$$U = \text{span}\{(1, 1, 0), (0, 1, 1)\},$$

ao longo do subespaço  $V = \text{span}\{(1, 0, 1)\}$ .

- 7.5. Encontre a projeção  $P : \mathbb{R}^4 \rightarrow \mathbb{R}^4$  de  $\mathbb{R}^4$  sobre o subespaço

$$U = \text{span}\{(1, 1, 0, 0), (0, 0, 1, 1)\},$$

ao longo do subespaço

$$V = \text{span}\{(1, 0, 1, 0), (1, 0, 0, 1)\}.$$

- 7.6. Encontre a projeção  $P : \mathbb{R}^4 \rightarrow \mathbb{R}^4$  de  $\mathbb{R}^4$  no plano  $\{(x, y, z, w); x = y, w = 0\}$ , ao longo do plano  $\{(x, y, z, w); x = 0, w = z\}$ . Encontre, também, a projeção complementar  $Q = I - P$ .
- 7.7. Considere o espaço vetorial  $\mathcal{P}_3$  dos polinômios com coeficientes reais de grau no máximo três. Mostre que  $\mathcal{P}_3$  é soma direta dos subespaços  $V = \{p \in \mathcal{P}_3, p'''(0) = p''(0) = 0\}$  e  $W = \{p \in \mathcal{P}_3; p''(0) = p'(0), p'''(0) = p(0)\}$ . Em seguida, encontre a projeção  $P : \mathcal{P}_3 \rightarrow \mathcal{P}_3$  de um polinômio  $p(x) = a_0 + a_1x + a_2x^2 + a_3x^3 \in \mathcal{P}_3$  sobre o subespaço  $V$ , ao longo do subespaço  $W$ .
- 7.8. Considere o espaço  $X = \mathcal{C}([-L, L], \mathbb{R})$  das funções reais contínuas definidas no intervalo  $[-L, L]$ , onde  $L > 0$ , e seja  $P$  uma função definida em  $X$  por

$$(P(f))(x) = \frac{f(x) + f(-x)}{2}, \quad \forall x \in [-L, L].$$

Mostre que  $P$  é um operador linear em  $X$  e que é uma projeção.

- 7.9. Sejam  $P, Q : X \rightarrow X$  dois operadores tais que  $I = P + Q$ . Mostre que  $P$  e  $Q$  são projeções se, e somente se,  $PQ = QP = 0$ .
- 7.10. Sejam  $P, Q : X \rightarrow X$  dois operadores tais que  $I = P + Q$ . Mostre que  $P$  e  $Q$  são projeções se, e somente se,  $\text{ker}(P) \cap \text{ker}(Q)$  e  $\text{Im}(P) \cap \text{Im}(Q)$  são triviais.

- 7.11.** Seja  $T \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$ . Suponha que o espaço seja soma direta  $X = U \oplus V$  de dois subespaços invariantes  $U$  e  $V$  de  $T$ . Suponha que  $P$  seja a projeção de  $X$  sobre  $U$  ao longo de  $V$ . Mostre que o operador e a projeção comutam entre si, i.e.  $TP = PT$ . (Dica: mostre primeiro que  $PTQ = 0$  e  $QTP = 0$ , onde  $Q = I - P$ ). (Essa resultado complementa o [Teorema 5.1](#), dando a comutatividade como caracterização de subespaços complementares invariantes. Veja, também, a [Observação 5.1](#))
- 7.12.** Complementando a [Observação 5.1](#), temos a seguinte caracterização. Seja  $T \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$ . Suponha que  $X = S_1 \oplus \dots \oplus S_m$ , onde  $S_j$  são subespaços de  $X$ , e seja  $P_j$  o operador projeção sobre  $S_j$ , ao longo de  $\bigoplus_{i \neq j} S_i$ . Então  $S_1, \dots, S_m$  são todos invariantes se, e somente se,  $TP_j = P_jT$ , para todo  $j = 1, \dots, m$ .
- 7.13.** Seja  $P : X \rightarrow X$  um operador projeção, em um espaço vetorial  $X$ . Defina  $R = I - 2P$ . Mostre que  $R$  é involutivo.
- 7.14.** Seja  $R : X \rightarrow X$  um operador linear em um espaço vetorial  $X$ . Mostre que  $R$  é involutivo se, e somente se, existem projeções  $P, Q : X \rightarrow X$  tais que  $P + Q = I$  e  $R = P - Q$ .
- 7.15.** Verifique a validade do [Teorema 5.2](#).
- 7.16.** Demostre o resultado descrito no [Exercício 5.41](#), da independência da fração em relação à base.

## CAPÍTULO 5

### Operadores Nilpotentes

Operadores nilpotentes têm um papel fundamental na teoria espectral de operadores lineares. São parte essencial da forma normal de Jordan de um operador linear em dimensão finita, discutida no [Capítulo 6](#). Operadores com estruturas especiais associadas ao produto interno, como operadores *normais*, que incluem os autoadjuntos e outros que aparecem com frequência em aplicações, prescindem desses operadores nilpotentes, mas, no caso geral, esses operadores são essenciais.

#### 1. Fundamentos

**1.1. Definição.** Um operador nilpotente é, por definição, um operador cuja alguma potência inteira positiva o anula completamente.

DEFINIÇÃO 1.1. *Seja  $X$  um espaço vetorial. Um operador  $N \in \mathcal{L}(X)$  é dito **nilpotente** quando existe um inteiro positivo  $k \in \mathbb{N}$  tal que  $N^k = 0$ . O **índice de nilpotência** de  $N$  é o menor inteiro  $k \in \mathbb{N}$  com essa propriedade.*

Ou seja, aplicando um operador nilpotente  $N$  sucessivamente, o espaço é, eventualmente, todo colapsado na origem:

$$X \mapsto N(X) \mapsto N^2(X) \mapsto \cdots \mapsto N^{k-1}(X) \mapsto \{0\}.$$

#### 1.2. Exemplos.

EXEMPLO 1.1. Em  $\mathbb{R}^2$ , o exemplo clássico é dado pelo operador

$$N(x, y) = (y, 0),$$

que é representado, na base canônica, pela matriz

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

O eixo  $x$  é levado na origem, enquanto que o plano todo  $xy$  é levado primeiro no eixo  $x$ , que em seguida é levado na origem.

EXEMPLO 1.2. O exemplo clássico de operador nilpotente em  $\mathbb{R}^n$  é o operador

$$\mathbf{N}(x_1, x_2, x_3, \dots, x_n) = (x_2, x_3, \dots, x_n, 0).$$

O espaço diminui de uma dimensão a cada aplicação do operador:

$$\mathbb{R}^n \xrightarrow{\mathbf{N}} \mathbb{R}^{n-1} \times \{0\} \xrightarrow{\mathbf{N}} \mathbb{R}^{n-2} \times \{(0, 0)\} \xrightarrow{\mathbf{N}} \dots \xrightarrow{\mathbf{N}} \{0\}.$$

EXEMPLO 1.3. Um exemplo *não-canônico* é dado pelo operador  $\mathbf{N} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  dado por

$$\mathbf{N}(x, y) = (x - y, x - y), \quad (x, y) \in \mathbb{R}^2.$$

Este operador leva o plano todo na reta  $y = x$ , enquanto que a reta  $y = x$  é levada na origem. Temos

$$\begin{aligned} \text{Im}(\mathbf{N}^2) &= \mathbf{N}^2(X) = \{0\}, \\ \text{Im}(\mathbf{N}) &= \mathbf{N}(X) = \{(x, y); y = x\}, \\ \text{Im}(\mathbf{N}^0) &= \text{Im}(\mathbf{I}) = \mathbb{R}^2. \end{aligned}$$

EXEMPLO 1.4. No espaço  $X = \mathcal{P}_n$  dos polinômios de grau no máximo  $n$ , o operador derivada  $\mathbf{D} : \mathcal{P}_n \rightarrow \mathcal{P}_n$ ,  $(\mathbf{D}p)(x) = p'(x)$ , para  $p \in \mathcal{P}_n$ , é um operador nilpotente com índice  $n + 1$ , que é a dimensão do espaço.

EXEMPLO 1.5. No espaço  $X = \mathcal{P}$  dos polinômios de ordem arbitrária, o operador derivada não é nilpotente, pois para cada potência  $n \in \mathbb{N}$ , temos  $x^n \in X$  com  $\mathbf{D}^{n+1}x^n = 0$  mas  $\mathbf{D}^n x^n = n!$ .

**1.3. Propriedades fundamentais.** Observe que cada  $\mathbf{N}^j(X) = \text{Im}(\mathbf{N}^j)$  da sequência abaixo é um subespaço vetorial que está no núcleo da potência  $\mathbf{N}^{k-j}$ . É possível mostrar que cada subespaço é um subespaço próprio do subespaço anterior, i.e.

$$X = \mathbf{N}^0(X) \supsetneq \mathbf{N}(X) \supsetneq \dots \supsetneq \mathbf{N}^{k-1}(X) \supsetneq \mathbf{N}^k(X) = \{0\}. \quad (1.1)$$

De maneira similar, cada núcleo  $\ker(\mathbf{N}^{k-j})$ ,  $j = 0, \dots, k$ , é subespaço próprio do núcleo seguinte. Mas a imagem  $\mathbf{N}^j(X) = \text{Im}(\mathbf{N}^j)$  pode não coincidir com o núcleo  $\ker(\mathbf{N}^{k-j})$ .

Vamos formalizar, a seguir, o resultado sobre o encadeamento próprio dos núcleos, mas o resultado sobre (1.1) deixamos como **Exercício 5.4**.

TEOREMA 1.1. *Seja  $\mathbf{N} : X \rightarrow X$  um operador nilpotente em um espaço vetorial  $X$ , com índice de nilpotência  $k \in \mathbb{N}$ . Então*

$$\{0\} \subsetneq \ker(\mathbf{N}) \subsetneq \dots \subsetneq \ker(\mathbf{N}^{k-1}) \subsetneq \ker(\mathbf{N}^k) = X, \quad (1.2)$$

*ou seja, cada  $\ker(\mathbf{N}^j)$  é subespaço próprio do nível  $\ker(\mathbf{N}^{j+1})$  imediatamente superior, para  $j = 0, \dots, k - 1$ .*

DEMONSTRAÇÃO. Primeiramente, se  $\ker(\mathbf{N})$  fosse trivial, então  $\mathbf{N}$  seria invertível e, com isso, todas as suas potências  $\mathbf{N}^k$  seriam invertíveis, com inversa  $(\mathbf{N}^{-1})^k$ . Assim, nenhuma potência  $\mathbf{N}^k$  seria nula e  $\mathbf{N}$  não poderia ser nilpotente. Logo, a primeira inclusão própria  $\{\mathbf{0}\} \subsetneq \ker(\mathbf{N})$  é verdadeira. Suponha, agora, que  $\ker(\mathbf{N}^j) = \ker(\mathbf{N}^{j+1})$  para algum  $1 \leq j \leq k-1$ . Nesse caso, para um  $\mathbf{u} \in X$  arbitrário, temos  $\mathbf{0} = \mathbf{N}^k \mathbf{u} = \mathbf{N}^{j+1} \mathbf{N}^{k-1-j} \mathbf{u}$ , de modo que  $\mathbf{N}^{k-1-j} \mathbf{u}$  está no núcleo de  $\mathbf{N}^{j+1}$ . Como esse núcleo é igual ao de  $\mathbf{N}^j$ , temos, também, que  $\mathbf{N}^{k-1} \mathbf{u} = \mathbf{N}^j \mathbf{N}^{k-1-j} \mathbf{u} = \mathbf{0}$ . Como  $\mathbf{u} \in X$  é arbitrário, isso mostra que  $\mathbf{N}^{k-1} = \mathbf{0}$ , contrariando a hipótese de que  $k$  é o menor índice que anula  $\mathbf{N}$ . Portanto, todas as inclusões devem ser próprias.  $\square$

COROLÁRIO 1.1. *Seja  $\mathbf{N} : X \rightarrow X$  um operador nilpotente em um espaço vetorial  $X$  de dimensão finita. Então  $\mathbf{N}$  não pode ter um grau de nilpotência maior do que a dimensão do espaço.*

DEMONSTRAÇÃO. Isso segue imediatamente de (1.2), visto que, a cada potência, a dimensão do núcleo aumenta de pelo menos uma unidade, não podendo passar da dimensão do espaço todo.  $\square$

## 2. Ciclos

**2.1. Cadeias de vetores e ciclos.** No caso de operadores nilpotentes, a cadeia de subespaços formada pela ação do operador linear no espaço  $X$  pode ser representada por uma ou mais cadeias formadas pela ação do operador em um ou mais vetores. Eventualmente, os vetores da cadeia são levados no vetor nulo, fechando um *ciclo*. Vamos formalizar essa noção. As definições abaixo não requerem que o operador seja nilpotente. Vamos fazer essas definições em um contexto mais geral. Isso será explorado mais pra frente, no [Capítulo 6](#), como partes de um quebra-cabeça mais complicado, válido para operadores quaisquer. No momento, no entanto, vamos usar isso, logo em seguida, para investigar os operadores nilpotentes.

DEFINIÇÃO 2.1. *Seja  $\mathbf{T} : X \rightarrow X$  um operador linear em um espaço vetorial  $X$ . Dado um certo vetor  $\mathbf{u} \in X$  e um inteiro não negativo  $j$ , a sequência  $\{\{\mathbf{u}, \mathbf{T}\mathbf{u}, \dots, \mathbf{T}^j \mathbf{u}\}\}$  é chamada de **cadeia** de vetores gerada por  $\mathbf{u}$ , de comprimento  $j+1$ . Esta cadeia é chamada de um **ciclo** quando é linearmente independente e maximal, no sentido de que  $\mathbf{T}^{j+1} \mathbf{u}$  é combinação linear das iterações anteriores, i.e.  $\{\{\mathbf{u}, \mathbf{T}\mathbf{u}, \dots, \mathbf{T}^{j+1} \mathbf{u}\}\}$  é linearmente dependente.*

Certos ciclos são longos o suficiente para formarem uma base. Nesse sentido, temos a seguinte definição.

DEFINIÇÃO 2.2. *Seja  $\mathbf{T} : X \rightarrow X$  um operador linear em um espaço vetorial  $X$  de dimensão finita  $n = \dim(X) \in \mathbb{N}$ . Quando existe um ciclo  $\alpha = \{\{\mathbf{u}, \mathbf{T}\mathbf{u}, \dots, \mathbf{T}^{n-1} \mathbf{u}\}\}$*

que forma uma base de  $X$ , dizemos que  $T$  é um **operador cíclico**, que  $\mathbf{u}$  é um **vetor cíclico** para  $T$  e que  $\mathcal{a}$  é o **ciclo** gerado por  $\mathbf{u}$ . A **base cíclica**  $\mathcal{b} = \{\{T^{n-1}\mathbf{u}, \dots, T\mathbf{u}, \mathbf{u}\}\}$  de  $X$ , para o operador  $T$ , é a base  $\mathcal{a}$  em ordem inversa.

EXEMPLO 2.1. Considere, por exemplo, o operador em  $\mathbb{R}^3$  associado à matriz

$$\mathbf{A} = \begin{bmatrix} 3 & -1 & 1 \\ -2 & 4 & 2 \\ -1 & 1 & 5 \end{bmatrix}$$

Então o vetor  $\mathbf{u} = (1, 1, 1)$  gera o ciclo

$$\{\{\mathbf{u}, \mathbf{A}\mathbf{u}, \mathbf{A}^2\mathbf{u}\}\} = \{(1, 1, 1), (3, 4, 5), (10, 20, 26)\},$$

que forma uma base cíclica de  $\mathbb{R}^3$ . Por outro lado, o vetor  $\mathbf{v} = (1, 0, -1)$  gera um ciclo mais curto, que não forma uma base,

$$\{\{\mathbf{v}, \mathbf{A}\mathbf{v}\}\} = \{(1, 0, -1), (2, -4, 6)\},$$

visto que

$$\mathbf{A}^2\mathbf{v} = (4, -32, -36) = -12(1, 0, -1) + 8(2, -4, 6) = -12\mathbf{v} + 8\mathbf{A}\mathbf{v}.$$

EXEMPLO 2.2. Um operador involutivo,  $R \in \mathcal{L}(X)$ ,  $R^2 = I$ , só pode gerar ciclos de, no máximo, dois vetores e, portanto, não admitem uma base cíclica em espaços de dimensão maior de que dois.

**2.2. Exemplos de ciclos em operadores nilpotentes.** Para ganharmos intuição, vamos ver exemplos de ciclos em operadores nilpotentes específicos. Observe que, nos exemplos abaixo, o ciclo, da **Definição 2.2**, termina quando  $T^{j+1}\mathbf{u} = \mathbf{0}$ .

EXEMPLO 2.3. Em  $\mathbb{R}^n$ , o operador canônico que é nilpotente com índice  $n$ , igual à dimensão, é o operador visto no **Exemplo 1.2**, definido por

$$\mathbf{N}(x_1, x_2, x_3, \dots, x_n) = (x_2, x_3, \dots, x_n, 0),$$

que, na base canônica  $e = \{\{e_1, \dots, e_n\}\}$ , tem a forma

$$[\mathbf{N}]_e = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

Podemos representar esse operador pela seguinte cadeia de vetores, exibindo a ação do operador na base:

$$\mathbf{e}_n \xrightarrow{\mathbf{N}} \mathbf{e}_{n-1} \xrightarrow{\mathbf{N}} \dots \xrightarrow{\mathbf{N}} \mathbf{e}_2 \xrightarrow{\mathbf{N}} \mathbf{e}_1 \xrightarrow{\mathbf{N}} \mathbf{0}.$$

Podemos visualizar isso da seguinte forma mais enfeitada, que é útil, principalmente, no caso de vários ciclos:



Vemos que esse operador é cíclico e que a base canônica é uma base cíclica.

Não custa ressaltar que esses são só vetores representativos da ação do operador no espaço. Todos os vetores múltiplos desses acompanham o ciclo, de modo que todo o espaço é “colapsado” em subespaços cada vez menores, até serem levados todos na origem.

OBSERVAÇÃO 2.1. A ordenação da base influencia, naturalmente, na forma da representação. No exemplo anterior, onde  $\mathbf{N}(x_1, \dots, x_n) = (x_2, x_3, \dots, x_n, 0)$ , ao ordenarmos os vetores da base canônica na ordem da cadeia de vetores,  $\mathcal{a} = \{\{\mathbf{e}_n, \dots, \mathbf{e}_1\}\}$ , obtemos

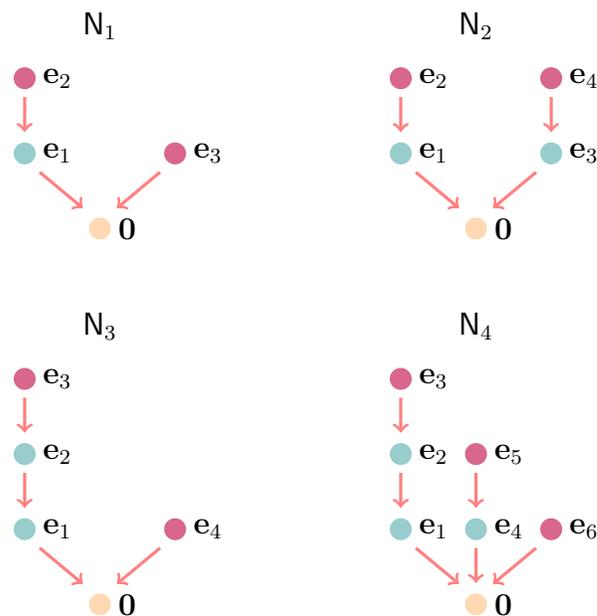
$$[\mathbf{N}]_{\mathcal{a}} = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}$$

A escolha da base cíclica ser na ordem inversa é apenas para definir a forma canônica como sendo aquela formada por 1's na superdiagonal. Apenas convenção.

EXEMPLO 2.4. Podemos ter operadores nilpotentes com um índice menor do que a dimensão do espaço. Por exemplo, considerando os operadores

$$\begin{aligned} \mathbf{N}_1(x, y, z) &= (y, 0, 0), \\ \mathbf{N}_2(x, y, z, w) &= (y, 0, w, 0), \\ \mathbf{N}_3(x, y, z, w) &= (y, z, 0, 0), \\ \mathbf{N}_4(x, y, z, u, v, w) &= (y, z, 0, v, 0, 0), \end{aligned}$$

temos  $N_1$  com índice  $k = 2$ , em  $\mathbb{R}^3$ ;  $N_2$  com índice  $k = 2$ , em  $\mathbb{R}^4$ ;  $N_3$  com índice  $k = 3$ , em  $\mathbb{R}^4$  e  $N_4$  com índice  $k = 3$  em  $\mathbb{R}^6$ . É ilustrativo olhar essas transformações de acordo com as suas ações nos elementos da base:



Podemos, ainda, olhar as representação matriciais desses operadores, na base canônica:

$$[N_1]_e = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad [N_2]_e = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix};$$

$$[N_3]_e = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}; \quad [N_4]_e = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

EXEMPLO 2.5. Todos os exemplos acima ilustram a ação do operador através da ação dele em cadeias de vetores da base canônica, por isso suas colunas e linhas só contém um único elemento não nulo. Mas vários outros operadores nilpotentes

existem de outra forma. Por exemplo, o operador considerado no **Exemplo 1.3**,

$$\mathbf{N}(x, y) = (x - y, x - y), \quad (x, y) \in \mathbb{R}^2,$$

leva o plano todo na reta  $y = x$ , enquanto que a reta  $y = x$  é levada na origem, sendo, portanto, o núcleo de  $\mathbf{N}$ . Tomando  $\mathbf{w}_1 = (1, 1)$  como a base do núcleo, podemos, em seguida, escolher qualquer vetor que é levado em  $\mathbf{w}_1$ , por exemplo,  $(1, 0)$  ou  $(0, 1)$ , ou qualquer vetor  $\mathbf{w}_2 \in (1, 0) + \ker(\mathbf{N}) = \{(1, 0) + (s, s); s \in \mathbb{R}\}$  que teremos

$$\mathbf{w}_2 \mapsto \mathbf{w}_1 = \mathbf{N}\mathbf{w}_2 \mapsto \mathbf{N}^2\mathbf{w}_2 = \mathbf{0}.$$

Esse operador é representado na base canônica por

$$[\mathbf{N}]_e = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}$$

enquanto que, na base cíclica  $\mathcal{b} = \{\{\mathbf{w}_1, \mathbf{w}_2\}\}$ , é simplesmente

$$[\mathbf{N}]_{\mathcal{b}} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

**EXEMPLO 2.6.** Mais geralmente, podemos formar qualquer ciclo, ou conjunto de ciclos, envolvendo todos os elementos de uma base, para construir um operador nilpotente. Por exemplo, a transformação  $\mathbf{N}$  com o ciclo

$$(1, 1, 1) \mapsto (2, 1, 0) \mapsto (3, 0, 0) \mapsto (0, 0, 0)$$

é dada por

$$\begin{aligned} \mathbf{N}(x, y, z) &= \mathbf{N}(z(1, 1, 1) + (y - z)(2, 1, 0) + (x - 2y + z)(3, 0, 0)) \\ &= z(2, 1, 0) + (y - z)(3, 0, 0) = (3y - z, z, 0), \end{aligned}$$

cuja representação na base canônica é

$$[\mathbf{N}]_e = \begin{bmatrix} 0 & 3 & -1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

**EXEMPLO 2.7.** Para uma matriz mais cheia, considere, por exemplo, os ciclos

$$(1, 1, 0, 4) \mapsto (3, 0, 2, 0) \mapsto (0, 1, 0, 1) \mapsto (0, 0, 0, 0)$$

e

$$(0, 5, 1, 0) \mapsto (0, 0, 0, 0).$$

A transformação associada a esse conjunto de ciclos, na representação canônica

$$(x, y, z, w) = x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3 + w\mathbf{e}_4,$$

é melhor expressa na forma matricial

$$[\mathbf{T}]_e = \begin{bmatrix} -123 & -37 & 185 & 40 \\ -3 & -1 & 5 & 1 \\ -80 & -24 & 120 & 26 \\ -12 & -4 & 20 & 4 \end{bmatrix}.$$

É claro que, na base

$$\mathcal{B} = \{(1, 1, 0, 4), (3, 0, 2, 0), (0, 1, 0, 1), (0, 5, 1, 0)\},$$

a transformação tem a forma

$$[\mathbf{T}]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

**OBSERVAÇÃO 2.2.** No [Exemplo 2.5](#), onde  $\mathbf{N}(x, y) = (x - y, x - y)$ , após selecionar  $\mathbf{w}_1 = (1, 1)$  como base do núcleo, tivemos a opção de escolher qualquer vetor que é levado em  $\mathbf{w}_1$ , ou seja, qualquer vetor  $\mathbf{w}_2 \in (1, 0) + \ker(\mathbf{N}) = \{(1, 0) + (s, s); s \in \mathbb{R}\}$ . Isso ilustra que a escolha da base que leva a um ciclo (ou conjunto de ciclos, em outras situações) não é única, visto que a escolha de  $\mathbf{w}_2$  é livre, dentro de um certo espaço afim. Além disso, poderíamos ter selecionado qualquer outro vetor como base do núcleo. Mas, nesse caso, qualquer escolha nos daria, de qualquer forma, um ciclo  $\mathbf{w}_2 \mapsto \mathbf{w}_1 \mapsto \mathbf{0}$ . Caso haja mais de um ciclo, podemos chegar em formas finais ligeiramente diferentes, reordenando os ciclos, mas o ponto principal dessa observação é a de que a escolha do ciclo não é única.

**OBSERVAÇÃO 2.3.** Ainda com base no [Exemplo 2.5](#), onde  $\mathbf{N}(x, y) = (x - y, x - y)$ , escolhamos primeiro um elemento  $\mathbf{w}_1$  em  $\ker(\mathbf{T})$  para em seguida escolher uma pré-imagem  $\mathbf{w}_2$  de  $\mathbf{w}_1$ . Mas isso só funcionou de maneira fácil porque o núcleo é unidimensional e não tivemos liberdade nessa escolha. Se o núcleo tiver dimensão maior, podemos acabar escolhendo uma base do núcleo que não tenha pré-imagem. Isso é melhor discutido no próximo [Exemplo 2.8](#).

**EXEMPLO 2.8.** Começar construindo o ciclo “por baixo”, como foi feito no [Exemplo 2.5](#) e discutido no [Observação 2.3](#), nem sempre funciona, ou requer cuidados maiores. Por exemplo, no operador  $\mathbf{N}(x, y, z) = (z, z, 0)$ , o núcleo é o plano  $xy$ ,  $\ker(\mathbf{N}) = \{z = 0\}$ . Se escolhermos  $\{\mathbf{e}_1, \mathbf{e}_2\}$  como base do núcleo, não vamos encontrar nenhum  $\mathbf{w}_3 \in X$  que seja levado em algum desses dois vetores do núcleo. Desse jeito, não vamos encontrar um ciclo caracterizando o operador. Temos a opção de escolher primeiro um vetor  $\mathbf{w}_1 \in \ker(\mathbf{T}) \cap \text{Im}(\mathbf{T})$ , completar a base de  $\ker(\mathbf{T})$  com

um vetor  $\mathbf{w}_2$  e por último encontrar  $\mathbf{w}_3$  que seja pré-imagem de  $\mathbf{w}_2$ . Mas esse procedimento pode se tornar bastante complexo em casos de vários ciclos. O mais fácil é começar “de cima”, com um  $\mathbf{w}_3$  em  $X \setminus \ker(\mathbf{N})$ , em seguida tomar  $\mathbf{w}_1 = \mathbf{N}\mathbf{w}_3$  e, por último, completar a base de  $\ker(\mathbf{N})$  com um  $\mathbf{w}_2$ . Esse é o procedimento que será generalizado.

**2.3. Caracterização de ciclos em operadores nilpotentes.** Bom, já vimos exemplos suficientes para termos uma intuição melhor sobre esses operadores nilpotentes. A chave são os ciclos. O tamanho dos ciclos está ligado ao índice de nilpotência. Podemos ter vários ciclos, pelo menos um do tamanho do índice de nilpotência, com outros podendo ser mais curtos. Essa é a ideia. Vamos ver, de fato, que qualquer operador nilpotente em dimensão finita é completamente definido por um ciclo ou conjunto de ciclos, sendo pelo menos um deles do tamanho do índice de nilpotência. Vamos tornar essas ideias mais precisas. Para isso, vamos começar com o caso em que o índice de nilpotência é igual à dimensão do espaço, caso este em que só pode haver um ciclo e o operador é, então, cíclico. Antes, vamos ver o seguinte resultado.

**PROPOSIÇÃO 2.1.** *Seja  $\mathbf{N} : X \rightarrow X$  um operador nilpotente em um espaço vetorial  $X$ . Se  $\mathbf{u} \in X$  é tal que  $\mathbf{N}^j \mathbf{u} = \mathbf{0}$  mas  $\mathbf{N}^{j-1} \mathbf{u} \neq \mathbf{0}$ , para algum  $j \in \mathbb{N}$ , então o conjunto ordenado*

$$\{\{\mathbf{u}, \mathbf{N}\mathbf{u}, \dots, \mathbf{N}^{j-1} \mathbf{u}\}\}$$

*é linearmente independente e, portanto, forma um ciclo.*

**DEMONSTRAÇÃO.** Suponha que

$$\lambda_1 \mathbf{N}^{j-1} \mathbf{u} + \lambda_2 \mathbf{N}^{j-2} \mathbf{u} + \dots + \lambda_{j-1} \mathbf{N}\mathbf{u} + \lambda_j \mathbf{u} = \mathbf{0}.$$

Precisamos mostrar que todos os escalares se anulam. Aplicamos  $\mathbf{N}^{j-1}$ , de modo que

$$\mathbf{N}^{j-1} (\lambda_1 \mathbf{N}^{j-1} \mathbf{u} + \lambda_2 \mathbf{N}^{j-2} \mathbf{u} + \dots + \lambda_{j-1} \mathbf{N}\mathbf{u} + \lambda_j \mathbf{u}) = \mathbf{N}^{j-1} \mathbf{0} = \mathbf{0}.$$

Por outro lado, usando que  $\mathbf{N}^{j'} \mathbf{u} = \mathbf{N}^{j'-j} \mathbf{N}^j \mathbf{u} = \mathbf{N}^{j'-j} \mathbf{0} = \mathbf{0}$  para todo  $j' \geq j$ , obtemos

$$\begin{aligned} & \mathbf{N}^{j-1} (\lambda_1 \mathbf{N}^{j-1} \mathbf{u} + \lambda_2 \mathbf{N}^{j-2} \mathbf{u} + \dots + \lambda_{j-1} \mathbf{N}\mathbf{u} + \lambda_j \mathbf{u}) \\ &= \lambda_1 \mathbf{N}^{j-1} \mathbf{N}^{j-1} \mathbf{u} + \lambda_2 \mathbf{N}^{j-1} \mathbf{N}^{j-2} \mathbf{u} + \dots + \lambda_{j-1} \mathbf{N}^{j-1} \mathbf{N}\mathbf{u} + \lambda_j \mathbf{N}^{j-1} \mathbf{u} \\ &= \lambda_1 \mathbf{N}^{2j-2} \mathbf{u} + \lambda_2 \mathbf{N}^{2j-3} \mathbf{u} + \dots + \lambda_{j-1} \mathbf{N}^{j-2} \mathbf{N}\mathbf{u} + \lambda_j \mathbf{N}^{j-1} \mathbf{u} \\ &= \lambda_j \mathbf{N}^{j-1} \mathbf{u}. \end{aligned}$$

Logo,

$$\lambda_j \mathbf{N}^{j-1} \mathbf{u} = \mathbf{0}.$$

Como  $\mathbf{N}^{j-1} \mathbf{u} \neq \mathbf{0}$ , segue que  $\lambda_j = 0$ . Sobra, então,

$$\lambda_1 \mathbf{N}^{j-1} \mathbf{u} + \lambda_2 \mathbf{N}^{j-2} \mathbf{u} + \dots + \lambda_{j-1} \mathbf{N}\mathbf{u} = \mathbf{0}.$$

Aplicando agora  $\mathbf{N}^{j-2}$  e proseguindo de maneira análoga, chegamos a

$$\lambda_{j-1} \mathbf{N}^{j-1} \mathbf{u} = \mathbf{0}.$$

Novamente, obtemos que  $\lambda_{j-1} = 0$ . Repetimos o procedimento até chegar em

$$\lambda_1 \mathbf{N}^{j-1} \mathbf{u} = \mathbf{0}.$$

Como antes, isso implica em  $\lambda_1 = 0$ . Portanto, todos os coeficientes se anulam, i.e.  $\lambda_1 = \dots = \lambda_j = 0$ , mostrando que o conjunto é linearmente independente.  $\square$

### 3. Forma normal de operados nilpotentes

**3.1. Forma normal no caso operadores nilpotentes cíclicos.** Com base no que já vimos, podemos caracterizar, nesse momento, os operadores nilpotentes com índice de nilpotência igual à dimensão do espaço, que nos dão operadores cíclicos.

**TEOREMA 3.1.** *Seja  $\mathbf{N} : X \rightarrow X$  um operador nilpotente em um espaço de dimensão finita, com grau de nilpotência igual à dimensão do espaço,  $n = \dim(X) \in \mathbb{N}$ . Então, para  $\mathbf{u} \in \ker(\mathbf{N}^n) \setminus \ker(\mathbf{N}^{n-1})$  arbitrário, temos que*

$$\mathcal{b} = \{\{\mathbf{N}^{n-1} \mathbf{u}, \dots, \mathbf{N} \mathbf{u}, \mathbf{u}\}\}$$

é uma base de  $X$  para a qual  $\mathbf{N}$  é definido pelo ciclo

$$\mathbf{u} \xrightarrow{\mathbf{N}} \mathbf{N} \mathbf{u} \cdots \xrightarrow{\mathbf{N}} \mathbf{N}^{n-1} \mathbf{u} \xrightarrow{\mathbf{N}} \mathbf{0}. \quad (3.1)$$

Assim, vemos que  $\mathbf{N}$  é um operador cíclico. Além disso, na base  $\mathcal{b}$ , temos a representação canônica

$$[\mathbf{N}]_{\mathcal{b}} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}. \quad (3.2)$$

**DEMONSTRAÇÃO.** Como  $\mathbf{N}$  tem grau de nilpotência  $n$ , então  $\mathbf{N}^{n-1} \neq \mathbf{0}$ . Com isso, existe  $\mathbf{u} \in X$  tal que  $\mathbf{N}^{n-1} \mathbf{u} \neq \mathbf{0}$ , com  $\mathbf{N}^n \mathbf{u} = \mathbf{0}$ . Graças à **Proposição 2.1**, o conjunto

$$\mathcal{b} = \{\{\mathbf{N}^{n-1} \mathbf{u}, \dots, \mathbf{N} \mathbf{u}, \mathbf{u}\}\}$$

é linearmente independente. Como  $\mathcal{b}$  contém  $n$  vetores e  $n$  é a dimensão do espaço  $X$ , então  $\mathcal{b}$  é uma base para  $X$ , provando o primeiro resultado. Naturalmente, essa base tem a propriedade cíclica (3.1). Como é uma base, a transformação  $\mathbf{N}$  está completamente determinada por esse ciclo. Por conta da ordenação escolhida para o ciclo, segue a representação (3.2).  $\square$

OBSERVAÇÃO 3.1. No caso em que o índice é menor do que a dimensão do espaço, temos, necessariamente, dois ou mais ciclos. Ciclos podem existir mesmo para operadores que não sejam nilpotentes, pois podem ser nilpotentes apenas em um determinado subespaço.

**3.2. Forma normal no caso geral.** Para provar o caso geral, de que um operador nilpotente possui uma base formada por um ou mais ciclos, vamos usar o **Teorema 1.1** e construir a base cíclica passo a passo.

Vamos, precisar, também, do seguinte resultado que garante que a união de ciclos formados por vetores linearmente independentes em um espaço complementar ao núcleo de uma dada potência do operador forma um conjunto linearmente independente.

PROPOSIÇÃO 3.1. *Seja  $N : X \rightarrow X$  um operador linear em um espaço vetorial  $X$ . Sejam  $\mathbf{w}_1, \dots, \mathbf{w}_m$ ,  $m \in \mathbb{N}$ , vetores em  $\ker(N^k)$ , para algum  $k \in \mathbb{N}$ , e suponha que eles sejam linearmente independentes e que o subespaço gerado por eles tenha interseção trivial com o núcleo  $\ker(N^{k-1})$ , i.e.*

$$\text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_m\} \cap \ker(N^{k-1}) = \{\mathbf{0}\}.$$

Então o conjunto de ciclos

$$\mathcal{a} = \{\{\mathbf{w}_1, \dots, N^{k-1}\mathbf{w}_1, \mathbf{w}_2, \dots, N^{k-1}\mathbf{w}_2, \dots, \mathbf{w}_m, \dots, N^{k-1}\mathbf{w}_m\}\}$$

é linearmente independente.

DEMONSTRAÇÃO. Primeiramente, como  $\mathbf{w}_1, \dots, \mathbf{w}_m \in \ker(N^k)$ , então

$$N^j \mathbf{w}_1, \dots, N^j \mathbf{w}_m \in \ker(N^{k-j}),$$

para todo  $j = 0, \dots, k-1$  (na verdade até  $k$ , mas isso não é importante). Além disso, da hipótese de interseção trivial, também temos que

$$\text{span}\{N^j \mathbf{w}_1, \dots, N^j \mathbf{w}_m\} \cap \ker(N^{k-j-1}) = \{\mathbf{0}\}.$$

De fato, se

$$\lambda_1 N^j \mathbf{w}_1 + \dots + \lambda_m N^j \mathbf{w}_m \in \ker(N^{k-j-1}),$$

então

$$N^{k-1} (\lambda_1 \mathbf{w}_1 + \dots + \lambda_m \mathbf{w}_m) = N^{k-j-1} (\lambda_1 N^j \mathbf{w}_1 + \dots + \lambda_m N^j \mathbf{w}_m) = \mathbf{0},$$

de modo que

$$\lambda_1 \mathbf{w}_1 + \dots + \lambda_m \mathbf{w}_m \in \ker(N^{k-1}),$$

além de, naturalmente,

$$\lambda_1 \mathbf{w}_1 + \dots + \lambda_m \mathbf{w}_m \in \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_m\}.$$

Como a interseção entre esses dois conjuntos é trivial, segue que

$$\lambda_1 \mathbf{w}_1 + \cdots + \lambda_m \mathbf{w}_m = \mathbf{0},$$

ou seja, a interseção entre  $\text{span}\{\mathbf{N}^j \mathbf{w}_1, \dots, \mathbf{N}^j \mathbf{w}_m\}$  e  $\ker(\mathbf{N}^{k-j-1})$  também é trivial.

Com isso, vamos mostrar que o conjunto  $\mathcal{a}$  é LI. Para tanto, suponha que

$$\begin{aligned} \lambda_{1,0} \mathbf{w}_1 + \cdots + \lambda_{1,k-1} \mathbf{N}^{k-1} \mathbf{w}_1 + \lambda_{2,0} \mathbf{w}_2 + \cdots + \lambda_{2,k-1} \mathbf{N}^{k-1} \mathbf{w}_2 \\ + \cdots + \lambda_{m,0} \mathbf{w}_m + \cdots + \lambda_{m,k-1} \mathbf{N}^{k-1} \mathbf{w}_m = \mathbf{0} \end{aligned}$$

Separamos os termos em grupos, de acordo com o nível em relação aos núcleos de potências do operador:

$$\mathbf{u}_j = \lambda_{1,j} \mathbf{N}^j \mathbf{w}_1 + \cdots + \lambda_{m,j} \mathbf{N}^j \mathbf{w}_m.$$

Para cada  $j$ , temos

$$\mathbf{u}_j \in \text{span}\{\mathbf{N}^j \mathbf{w}_1, \dots, \mathbf{N}^j \mathbf{w}_m\} \subset \ker(\mathbf{N}^{k-j}).$$

Começamos com  $j = 0$ , temos, por um lado,

$$\mathbf{u}_0 \in \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_m\},$$

enquanto que, por outro,

$$\mathbf{u}_0 = -\mathbf{u}_1 - \cdots - \mathbf{u}_{k-1} \in \ker(\mathbf{N}^{k-1}).$$

Como os subespaços têm interseção trivial, segue que

$$\mathbf{u}_0 = \mathbf{0}, \quad \mathbf{u}_1 + \cdots + \mathbf{u}_{k-1} = \mathbf{0}.$$

Agora, com

$$\mathbf{u}_0 = \lambda_{1,0} \mathbf{w}_1 + \cdots + \lambda_{m,0} \mathbf{w}_m = \mathbf{0},$$

e sendo os vetores  $\{\mathbf{w}_1, \dots, \mathbf{w}_m\}$  linearmente independentes, deduzimos que

$$\lambda_{1,0} = \cdots = \lambda_{m,0} = 0.$$

Suponha, então, por indução, que tenhamos mostrado que

$$\lambda_{1,i} = \cdots = \lambda_{m,i} = 0,$$

para todo  $i = 0, \dots, j-1$ , para algum  $1 \leq j \leq k-1$ . Vamos mostrar que o mesmo vale para  $j$ . Nesse ponto, sobraram

$$\mathbf{u}_j + \mathbf{u}_{j+1} + \cdots + \mathbf{u}_{k-1} = \mathbf{0}.$$

Como

$$\mathbf{u}_j \in \text{span}\{\mathbf{N}^j \mathbf{w}_1, \dots, \mathbf{N}^j \mathbf{w}_m\}$$

e

$$\mathbf{u}_{j+1} + \cdots + \mathbf{u}_{k-1} \in \ker(\mathbf{N}^{k-j-1}),$$

com

$$\text{span}\{\mathbf{N}^j \mathbf{w}_1, \dots, \mathbf{N}^j \mathbf{w}_m\} \cap \ker(\mathbf{N}^{k-j-1}) = \{\mathbf{0}\},$$

segue que

$$\mathbf{u}_j = \lambda_{1,j} \mathbf{N}^j \mathbf{w}_1 + \dots + \lambda_{m,j} \mathbf{N}^j \mathbf{w}_m = \mathbf{0},$$

ou seja,

$$\mathbf{N}^j (\lambda_{1,j} \mathbf{w}_1 + \dots + \lambda_{m,j} \mathbf{w}_m) = \mathbf{0},$$

Isso significa que

$$\lambda_{1,j} \mathbf{w}_1 + \dots + \lambda_{m,j} \mathbf{w}_m \in \ker(\mathbf{N}^j) \subset \ker(\mathbf{N}^{k-1}),$$

visto que  $j \leq k-1$ . Mas a interseção entre o  $\text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_m\}$  e  $\ker(\mathbf{N}^{k-1})$  é trivial. Lembrando que os vetores  $\mathbf{w}_1, \dots, \mathbf{w}_m$  formam um conjunto LI, deduzimos que

$$\lambda_{1,j} = \dots = \lambda_{m,j} = 0.$$

Isso completa a indução até  $j = k-1$ , mostrando que todos os escalares se anulam e o conjunto  $\mathcal{a}$  é LI.  $\square$

Na verdade, o resultado acima foi só um aquecimento para o resultado que vamos realmente precisar. Precisamos considerar o caso de ciclos de tamanhos variados. Vamos ver, então, a seguinte extensão do resultado acima.

**PROPOSIÇÃO 3.2.** *Seja  $\mathbf{N} : X \rightarrow X$  um operador linear em um espaço vetorial  $X$ . Sejam  $\mathbf{w}_1, \dots, \mathbf{w}_m$ ,  $m \in \mathbb{N}$ , vetores em  $X$  e  $k_1 \geq \dots \geq k_m \geq 1$  inteiros positivos tais que  $\mathbf{w}_j \in \ker(\mathbf{N}^{k_j}) \setminus \ker(\mathbf{N}^{k_j-1})$ . Suponha que para cada  $k_m \leq k \leq k_1$ , o conjunto “no mesmo nível”*

$$\mathcal{a}_k = \{\{\mathbf{N}^{k_j-k} \mathbf{w}_j; k \leq k_j \leq k_1\}\} \subset \ker(\mathbf{N}^k)$$

*seja linearmente independente e que o espaço gerado por  $\mathcal{a}_k$  tenha interseção trivial com  $\ker(\mathbf{N}^{k-1})$ , i.e.*

$$\text{span}(\mathcal{a}_k) \cap \ker(\mathbf{N}^{k-1}) = \{\mathbf{0}\}.$$

*Então o conjunto de ciclos*

$$\mathcal{a} = \{\{\mathbf{w}_1, \dots, \mathbf{N}^{k_1-1} \mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{N}^{k_2-1} \mathbf{w}_2, \dots, \mathbf{w}_m, \dots, \mathbf{N}^{k_m-1} \mathbf{w}_m\}\}$$

*é linearmente independente.*

**DEMONSTRAÇÃO.** Deixamos os detalhes da demonstração desse resultado para o leitor. É uma pequena variação da demonstração da **Proposição 3.1**. Para tal, é importante observar que

$$\mathbf{N}^k \mathcal{a}_k = \{\mathbf{N}^{k_j} \mathbf{w}_j; k \leq k_j \leq k_1\} = \{\mathbf{0}\}$$

e

$$\mathbf{0} \notin \mathbf{N}^{k-1} \mathfrak{a}_k = \{\mathbf{N}^{k_j-1} \mathbf{w}_j; k \leq k_j \leq k_1\},$$

ou seja,  $\mathfrak{a}_k$  está, de fato, “no nível” de  $\ker(\mathbf{N}^k)/\ker(\mathbf{N}^{k-1})$ . Assim, por conta da hipótese das interseções serem triviais, podemos separar, indutivamente, uma combinação linear

$$\begin{aligned} \lambda_{1,0} \mathbf{w}_1 + \cdots + \lambda_{1,k_1} \mathbf{N}^{k_1-1} \mathbf{w}_1 + \lambda_{2,0} \mathbf{w}_2 + \cdots + \lambda_{2,k_2} \mathbf{N}^{k_2-1} \mathbf{w}_2 + \cdots \\ + \lambda_{m,0} \mathbf{w}_m + \cdots + \lambda_{m,k_m} \mathbf{N}^{k_m-1} \mathbf{w}_m = \mathbf{0} \end{aligned}$$

em cada nível, obtendo

$$\sum_{j=1}^m \sum_{k=0, \dots, k_j-1} \lambda_{j,k} \mathbf{N}^k \mathbf{w}_j = \mathbf{0},$$

e em seguida usar a hipótese de que os vetores de cada nível são LI.  $\square$

Agora, podemos ir ao resultado geral de caracterização de operadores nilpotentes.

**TEOREMA 3.2.** *Seja  $\mathbf{N} : X \rightarrow X$  um operador nilpotente com índice de nilpotência  $k \in \mathbb{N}$ , em um espaço vetorial  $X$  de dimensão finita  $n = \dim(X) \in \mathbb{N}$ . Então existem vetores  $\mathbf{w}_1, \dots, \mathbf{w}_m$  e índices  $k_1, \dots, k_m \in \mathbb{N}$ , com  $m \in \mathbb{N}$ , tais cada*

$$\mathfrak{b}_j = \{\{\mathbf{N}^{k_j-1} \mathbf{w}_j, \dots, \mathbf{N} \mathbf{w}_j, \mathbf{w}_j\}\}$$

é um ciclo, i.e. é linearmente independente com  $\mathbf{N}^{k_j} \mathbf{w}_j = \mathbf{0}$ , e o conjunto

$$\mathfrak{b} = \mathfrak{b}_1 \cup \cdots \cup \mathfrak{b}_m$$

forma uma base de  $X$ . Além disso, na base  $\mathfrak{b}$ , o operador  $\mathbf{N}$  tem uma estrutura de blocos,

$$[\mathbf{N}]_{\mathfrak{b}} = \begin{bmatrix} \mathbf{J}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{J}_m \end{bmatrix}.$$

com blocos  $\mathbf{J}_j$  de tamanho  $k_j \times k_j$  e da forma

$$\mathbf{J}_j = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

DEMONSTRAÇÃO. Primeiramente, se  $k = 1$ , então  $\mathbf{N} = \mathbf{0}$  e basta escolhermos uma base qualquer  $\mathcal{B} = \{\{\mathbf{w}_1, \dots, \mathbf{w}_n\}\}$  de  $X$  e definir  $m = n$  e  $k_j = 1$ , para todo  $j = 1, \dots, n$ . Nesse caso,  $\mathcal{B}_j = \{\{\mathbf{w}_j\}\}$  e  $\mathbf{N}\mathbf{w}_j = \mathbf{0}$ , para todo  $j$ , concluindo a demonstração para esse caso. Os blocos nilpotentes são todos  $1 \times 1$ , com  $J_j = [0]$ .

Se  $k = n$ , basta aplicar o [Teorema 3.1](#) e obtermos um único bloco nilpotente  $J_1$  como em [\(3.2\)](#).

Falta considerar, então, o caso  $2 \leq k \leq n - 1$ . A ideia é escolher, sucessivamente, bases para os espaços quocientes  $\ker(\mathbf{N}^{k-j+1})/\ker(\mathbf{N}^{k-j})$ ,  $j = 1, \dots, k$ , e selecionar representantes para formar cada ciclo. Vamos aos detalhes.

Primeiramente, denotamos por  $\sim_{k-j}$  a relação de equivalência módulo  $\ker(\mathbf{N}^{k-j})$ ,  $j = 1, \dots, k$ , i.e.

$$\mathbf{u} \sim_{k-j} \mathbf{v} \Leftrightarrow \mathbf{v} - \mathbf{u} \in \ker(\mathbf{N}^{k-j}),$$

onde, por conveniência, incluímos  $\mathbf{N}^0 = \mathbf{I}$ , com  $\ker(\mathbf{N}^0) = \ker(\mathbf{I}) = \{\mathbf{0}\}$ . A classe de equivalência módulo  $\ker(\mathbf{N}^{k-j})$  de um elemento  $\mathbf{u}$  é denotada por  $[\mathbf{u}]_j$ .

Sejam

$$m_j = \dim(\ker(\mathbf{N}^{k-j+1})/\ker(\mathbf{N}^{k-j})).$$

Graças ao [Teorema 1.1](#), temos

$$m_j \geq 1, \quad \forall j = 1, \dots, k.$$

Além disso, usando a [Proposição 2.7](#), temos

$$\begin{aligned} n = \dim(X) &= \dim(\ker(\mathbf{N}^k)) \\ &= \dim(\ker(\mathbf{N}^k)/\ker(\mathbf{N}^{k-1})) + \dim(\ker(\mathbf{N}^{k-1})) \\ &= \dim(\ker(\mathbf{N}^k)/\ker(\mathbf{N}^{k-1})) + \dim(\ker(\mathbf{N}^{k-1})/\ker(\mathbf{N}^{k-2})) \\ &\quad + \dots + \dim(\ker(\mathbf{N}^2)/\ker(\mathbf{N})) + \dim(\ker(\mathbf{N})) \\ &= m_1 + m_2 + \dots + m_k. \end{aligned}$$

O objetivo, a partir desse ponto, é obter os representantes de bases apropriadas para os espaços quocientes acima.

Começamos, então, escolhendo representantes  $\mathbf{w}_1, \dots, \mathbf{w}_{m_1}$  para uma base do espaço quociente  $\ker(\mathbf{N}^k)/\ker(\mathbf{N}^{k-1})$ ,

$$\{\{[\mathbf{w}_1]_{k-1}, \dots, [\mathbf{w}_{m_1}]_{k-1}\}\}.$$

Em seguida, aplicando  $\mathbf{N}$ , “descemos” esses elementos para o nível  $\ker(\mathbf{N}^{k-1})/\ker(\mathbf{N}^{k-2})$ , obtendo os elementos

$$\{\{[\mathbf{N}\mathbf{w}_1]_{k-2}, \dots, [\mathbf{N}\mathbf{w}_{m_1}]_{k-2}\}\}. \quad (3.3)$$

Um fato importante é que esses elementos são linearmente independentes. Isso sai da [Proposição 3.1](#), mas uma demonstração mais direta é discutida na [Observação 3.4](#). Observe que isso nos dá que

$$m_2 \geq m_1.$$

Caso  $m_2 > m_1$ , podemos completar até uma base de  $\ker(\mathbf{N}^{k-1})/\ker(\mathbf{N}^{k-2})$ ,

$$\left\{ \left[ \left[ \mathbf{N}\mathbf{w}_1 \right] \right]_{k-2}, \dots, \left[ \left[ \mathbf{N}\mathbf{w}_{m_1} \right] \right]_{k-2}, \left[ \left[ \mathbf{w}_{m_1+1} \right] \right]_{k-2}, \dots, \left[ \left[ \mathbf{w}_{m_2} \right] \right]_{k-2} \right\},$$

com novos representantes  $\mathbf{w}_{m_1+1}, \dots, \mathbf{w}_{m_2}$ , com o entendimento de que esse conjunto extra é vazio, quando  $m_2 = m_1$ .

Suponhamos, por indução, que tenhamos encontrado bases de cada espaço quociente  $\ker(\mathbf{N}^{k-i+1})/\ker(\mathbf{N}^{k-i})$ , para  $i = 1, \dots, j < k$ , da forma

$$\left\{ \left[ \left[ \mathbf{N}^{i-1}\mathbf{w}_1 \right] \right]_{k-i}, \dots, \left[ \left[ \mathbf{N}^{i-1}\mathbf{w}_{m_1} \right] \right]_{k-i}, \dots, \left[ \left[ \mathbf{w}_{m_{i-1}+1} \right] \right]_{k-i}, \dots, \left[ \left[ \mathbf{w}_{m_i} \right] \right]_{k-i} \right\}.$$

Vamos “descer” mais um nível e completar até uma base de  $\ker(\mathbf{N}^{k-j})/\ker(\mathbf{N}^{k-j-1})$ , se necessário. A questão mais delicada é mostrar que, aplicando  $\mathbf{N}$  no caso  $i = j$ , o conjunto

$$\left\{ \left[ \left[ \mathbf{N}^j\mathbf{w}_1 \right] \right]_{k-j-1}, \dots, \left[ \left[ \mathbf{N}^j\mathbf{w}_{m_1} \right] \right]_{k-j-1}, \dots, \left[ \left[ \mathbf{N}\mathbf{w}_{m_{j-1}+1} \right] \right]_{k-j-1}, \dots, \left[ \left[ \mathbf{N}\mathbf{w}_{m_j} \right] \right]_{k-j-1} \right\} \quad (3.4)$$

é linearmente independente. Isso sai, agora, da [Proposição 3.2](#). Mas, novamente, podemos dar uma demonstração mais direta, como discutido na [Observação 3.5](#). Com isso, podemos completar até uma base de  $\ker(\mathbf{N}^{k-j})/\ker(\mathbf{N}^{k-j-1})$ ,

$$\left\{ \left[ \left[ \mathbf{N}^j\mathbf{w}_1 \right] \right]_{k-j-1}, \dots, \left[ \left[ \mathbf{N}^j\mathbf{w}_{m_1} \right] \right]_{k-j-1}, \dots, \left[ \left[ \mathbf{N}\mathbf{w}_{m_{j-1}+1} \right] \right]_{k-j-1}, \dots, \right. \\ \left. \left[ \left[ \mathbf{N}\mathbf{w}_{m_j} \right] \right]_{k-j-1}, \left[ \left[ \mathbf{w}_{m_{j+1}} \right] \right]_{k-j-1}, \dots, \left[ \left[ \mathbf{w}_{m_{j+1}} \right] \right]_{k-j-1} \right\}.$$

Assim, podemos concluir a indução até  $j + 1 = k$ .

Nesse momento, coletamos os seguintes vetores,

$$\mathcal{B}' = \left\{ \left\{ \mathbf{N}^{k-1}\mathbf{w}_1, \dots, \mathbf{N}^{k-1}\mathbf{w}_{m_1}, \mathbf{N}^{k-2}\mathbf{w}_1, \dots, \mathbf{N}^{k-2}\mathbf{w}_{m_2}, \dots, \mathbf{w}_1, \dots, \mathbf{w}_{m_k} \right\} \right\},$$

onde cada grupo

$$\mathcal{B}'_j = \left\{ \left\{ \mathbf{N}^{j-1}\mathbf{w}_1, \dots, \mathbf{N}^{j-1}\mathbf{w}_{m_1}, \dots, \mathbf{w}_{m_{j-1}+1}, \dots, \mathbf{w}_{m_j} \right\} \right\}$$

é linearmente independente. Precisamos mostrar, ainda, que o conjunto todo  $\mathcal{B}'$  é linearmente independente. Isso segue da [Proposição 3.2](#).

Obtemos, então, que o conjunto  $\mathcal{b}'$  de  $n = m_1 + m_2 + \dots + m_k$  vetores é linearmente independente, no espaço  $X$  de dimensão também  $n$ . Concluimos, assim, que  $\mathcal{b}'$  é base de  $X$ .

Para a base desejada  $\mathcal{b}$ , do enunciado do teorema, precisamos, apenas, reordenar a base  $\mathcal{b}'$ , separando cada ciclo

$$\mathcal{b}_j = \{ \{ \mathbf{N}^{k_j-1} \mathbf{w}_j, \dots, \mathbf{N} \mathbf{w}_j, \mathbf{w}_j \} \},$$

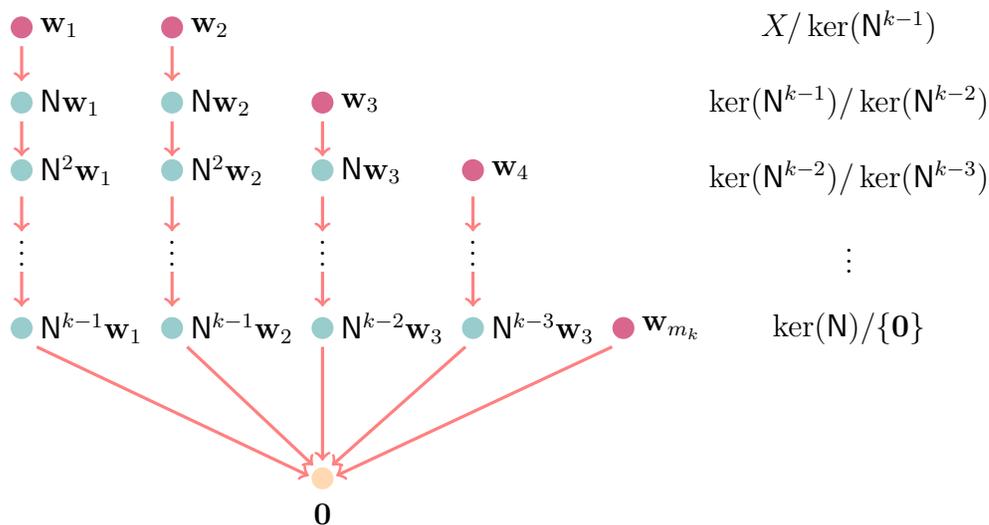
onde  $1 \leq k_j \leq k$  é dado por  $k$ , se  $1 \leq j \leq m_1$  e por  $k - i$ , se  $m_{i-1} + 1 \leq j \leq m_i$ , para  $i = 2, \dots, k$ .

Observe que cada nível  $\ker(\mathbf{N}^{k-j}) / \ker(\mathbf{N}^{k-j-1})$  tem dimensão  $m_j$ , mas  $m_{j-1}$  elementos vêm do nível de cima, com a base sendo acrescida de  $m_j - m_{j-1}$  vetores, quando necessário. Apenas esses novos elementos formam um novo ciclo. Assim, temos um total de

$$m = m_1 + (m_2 - m_1) + \dots + (m_k - m_{k-1}) = m_k$$

ciclos. Portanto, o número total  $m$  de ciclos é exatamente  $m_k$ . Escrevendo assim, temos que  $\mathcal{b} = \mathcal{b}_1 \cup \dots \cup \mathcal{b}_{m_k}$  é a base desejada.  $\square$

**OBSERVAÇÃO 3.2.** A demonstração do **Teorema 3.2** é construtiva. Escolhemos representantes para a base de  $X / \ker(\mathbf{N}^{k-1})$ . Descemos um nível, aplicando  $\mathbf{N}$ , e completamos com os representantes até uma base de  $\ker(\mathbf{N}^{k-1}) / \ker(\mathbf{N}^{k-2})$ , e assim sucessivamente, até uma base de  $\ker(\mathbf{N}) \equiv \ker(\mathbf{N}) / \{ \mathbf{0} \}$ . Esse processo de construção da base  $\mathcal{b}$  pode ser ilustrado pela seguinte figura:



OBSERVAÇÃO 3.3. Observe que cada  $\ker(\mathbf{N}^{k-j+1})/\ker(\mathbf{N}^{k-j})$  é isomorfo a qualquer complementar de  $\ker(\mathbf{N}^{k-j})$  em  $\ker(\mathbf{N}^{k-j+1})$ , ou seja a qualquer subespaço  $S' \subset \ker(\mathbf{N}^{k-j+1})$  tal que

$$\ker(\mathbf{N}^{k-j+1}) = \ker(\mathbf{N}^{k-j}) \oplus S'.$$

Também podemos dizer que  $m_j = \dim(\ker(\mathbf{N}^{k-j+1})/\ker(\mathbf{N}^{k-j}))$  é, também, a *codimensão* de  $\ker(\mathbf{N}^{k-j})$  como subespaço de  $\ker(\mathbf{N}^{k-j+1})$ .

**3.3. Mais exemplos.** Já vimos vários exemplos simples, na [Seção 2.2](#). Vamos ver mais alguns exemplos, a partir do processo descrito na [Observação 3.7](#).

EXEMPLO 3.1. Considere o operador  $\mathbf{N} : \mathbb{R}^6 \rightarrow \mathbb{R}^6$  dado por

$$\begin{aligned} \mathbb{T}(x, y, z, u, v, w) = (x - y, x - z, x - z, x - z + u - v, \\ x - z + u - w, x - z + u - w) \end{aligned}$$

que está associado à matriz

$$\mathbf{N} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 1 & -1 & 0 \\ 1 & 0 & -1 & 1 & 0 & -1 \\ 1 & 0 & -1 & 1 & 0 & -1 \end{bmatrix}$$

Calculamos

$$\mathbf{N}^2 = \begin{bmatrix} 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & -1 & 1 \\ 0 & -1 & 1 & 0 & -1 & 1 \\ 0 & -1 & 1 & 0 & -1 & 1 \end{bmatrix} \quad \text{e} \quad \mathbf{N}^3 = \mathbf{0}.$$

Portanto,

$$\begin{aligned} \ker(\mathbf{N}^3) &= \mathbb{R}^6 && (\text{dimensão } 6), \\ \ker(\mathbf{N}^2) &= \{(x, y, z, u, v, w); z = y, w = v\} && (\text{dimensão } 4), \\ \ker(\mathbf{N}) &= \{(x, y, z, u, v, w); z = y = x, w = v = u\} && (\text{dimensão } 2). \end{aligned}$$

Temos, também,

$$m_1 = \dim(\ker(\mathbf{N}^3)/\ker(\mathbf{N}^2)) = 6 - 4 = 2,$$

$$m_2 = \dim(\ker(\mathbf{N}^2)/\ker(\mathbf{N}^1)) = 4 - 2 = 2,$$

$$m_3 = \dim(\ker(\mathbf{N})/\ker(\mathbf{N}^0)) = 2 - 0 = 2.$$

Como representantes para uma base de  $\dim(\ker(\mathbf{N}^3)/\ker(\mathbf{N}^2))$  (ou, de outra forma, como geradores de um espaço complementar a  $\ker(\mathbf{N}^2)$ ), tomamos

$$\mathbf{w}_1 = (0, 1, -1, 0, 0, 0), \quad \mathbf{w}_2 = (0, 0, 0, 0, 1, -1) \in \ker(\mathbf{N}^3) \setminus \ker(\mathbf{N}^2).$$

A partir desses, temos

$$\mathbf{N}(\mathbf{w}_1) = (-1, 1, 1, 1, 1, 1), \quad \mathbf{N}(\mathbf{w}_2) = (0, 0, 0, -1, 1, 1).$$

Esses geram classes de equivalência linearmente independentes em  $\ker(\mathbf{N}^2)/\ker(\mathbf{N})$ , que tem, justamente, dimensão  $m_2 = 2$ . Em seguida, temos

$$\mathbf{N}^2(\mathbf{w}_1) = (-2, -2, -2, -2, -2, -2), \quad \mathbf{N}^2(\mathbf{w}_2) = (0, 0, 0, -2, -2, -2).$$

que geram classes de equivalência linearmente independentes em  $\ker(\mathbf{N})/\ker(\mathbf{N}^0)$ , com dimensão exatamente  $m_3 = 2$ . Assim, encontramos os ciclos

$$\begin{aligned} (0, 1, -1, 0, 0, 0) &\mapsto (-1, 1, 1, 1, 1, 1) \mapsto (-2, -2, -2, -2, -2, -2) \mapsto \mathbf{0}, \\ (0, 0, 0, 0, 1, -1) &\mapsto (0, 0, 0, -1, 1, 1) \mapsto (0, 0, 0, -2, -2, -2) \mapsto \mathbf{0}. \end{aligned}$$

Na base

$$\begin{aligned} \mathcal{B} &= \{\mathbf{N}^2\mathbf{w}_1, \mathbf{N}\mathbf{w}_1, \mathbf{w}_1, \mathbf{N}^2\mathbf{w}_2, \mathbf{N}\mathbf{w}_2, \mathbf{w}_2\} \\ &= \{(-2, -2, -2, -2, -2, -2), (-1, 1, 1, 1, 1, 1), (0, 1, -1, 0, 0, 0) \\ &\quad (0, 0, 0, -2, -2, -2), (0, 0, 0, -1, 1, 1), (0, 0, 0, 0, 1, -1)\}, \end{aligned}$$

temos a representação

$$[\mathbf{N}]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

EXEMPLO 3.2. O Exemplo 3.1 começou com um espaço quociente de dimensão dois e não foi necessário acrescentar mais vetores. Vamos fazer um outro exemplo,

em que isso é necessário. Vamos considerar o operador  $\mathbf{N} : \mathbb{R}^6 \rightarrow \mathbb{R}^6$  dado por

$$\mathbf{N}(x, y, z, u, v, w) = (x - 2y - z, 0, x - z, y + z + u - v/2 - w/2, \\ x + y + u - v/2 - w/2, x + y + u - v/2 - w/2)$$

que está associado à matriz

$$\mathbf{N} = \begin{bmatrix} 1 & -2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & -1/2 & -1/2 \\ 1 & 1 & 0 & 1 & -1/2 & -1/2 \\ 1 & 1 & 0 & 1 & -1/2 & -1/2 \end{bmatrix}$$

Calculamos

$$\mathbf{N}^2 = \begin{bmatrix} 0 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{e} \quad \mathbf{N}^3 = \mathbf{0}.$$

Portanto,

$$\ker(\mathbf{N}^3) = \mathbb{R}^6 \quad (\text{dimensão } 6),$$

$$\ker(\mathbf{N}^2) = \{(x, y, z, u, v, w); y = 0\} \quad (\text{dimensão } 5),$$

$$\ker(\mathbf{N}) = \{(x, y, z, u, v, w); y = 0; z = x = -u + v/2 + w/2\} \quad (\text{dimensão } 3).$$

Temos, assim,

$$m_1 = \dim(\ker(\mathbf{N}^3)/\ker(\mathbf{N}^2)) = 6 - 5 = 1,$$

$$m_2 = \dim(\ker(\mathbf{N}^2)/\ker(\mathbf{N}^1)) = 5 - 3 = 2,$$

$$m_3 = \dim(\ker(\mathbf{N})/\ker(\mathbf{N}^0)) = 3 - 0 = 3.$$

Como representante para uma base do espaço unidimensional  $\dim(\ker(\mathbf{N}^3)/\ker(\mathbf{N}^2))$  (ou, de outra forma, como gerador de um espaço complementar a  $\ker(\mathbf{N}^2)$ ), tomamos

$$\mathbf{w}_1 = (0, 1, 0, 0, 0, 0) \in \ker(\mathbf{N}^3) \setminus \ker(\mathbf{N}^2).$$

A partir desse, temos

$$\mathbf{N}(\mathbf{w}_1) = (-2, 0, 0, 1, 1, 1).$$

Para completar até uma base de  $\ker(\mathbf{N}^2)/\ker(\mathbf{N})$ , que tem dimensão  $m_2 = 2$ , acrescentamos o representante

$$\mathbf{w}_2 = (0, 0, 1, 0, 0, 0).$$

Observe que, se

$$\mathbf{u} \in \text{span}\{\mathbf{N}\mathbf{w}_1, \mathbf{w}_2\} \cap \ker(\mathbf{N}) = \{\mathbf{0}\},$$

então

$$\mathbf{u} = \alpha\mathbf{N}(\mathbf{w}_1) + \beta\mathbf{w}_2 = (-2\alpha, 0, \beta, \alpha, \alpha, \alpha) \in \ker(\mathbf{N}),$$

deve satisfazer a condição  $z = x = -u + v/2 + w/2$ , i.e.

$$\beta = -2\alpha = -\alpha + \alpha/2 + \alpha/2 = 0,$$

mostrando que  $\alpha = \beta = 0$ , ou seja,  $\mathbf{u} = \mathbf{0}$ , de modo que, como desejado,

$$\text{span}\{\mathbf{N}\mathbf{w}_1, \mathbf{w}_2\} \cap \ker(\mathbf{N}) = \{\mathbf{0}\}.$$

Descendo mais, obtemos

$$\mathbf{N}^2(\mathbf{w}_1) = (-2, 0, -2, 0, -2, -2), \quad \mathbf{N}(\mathbf{w}_2) = (-1, 0, -1, 1, 0, 0),$$

cujas classes de equivalência são linearmente independentes em  $\ker(\mathbf{N})/\ker(\mathbf{N}^0)$ , com dimensão  $m_3 = 3$ . Assim, precisamos de mais um representante para completar a base. Tomamos

$$\mathbf{w}_3 = (0, 0, 0, 0, 1, -1).$$

É fácil verificar que  $\mathbf{w}_3 \in \ker(\mathbf{N})$  e que  $\{\mathbf{N}^2(\mathbf{w}_1), \mathbf{N}(\mathbf{w}_2), \mathbf{w}_3\}$  é LI. Desta forma, encontramos os ciclos

$$(0, 1, 0, 0, 0, 0) \mapsto (-2, 0, 0, 1, 1, 1) \mapsto (-2, 0, -2, 0, -2, -2) \mapsto \mathbf{0},$$

$$(0, 0, 1, 0, 0, 0) \mapsto (-1, 0, -1, 1, 0, 0) \mapsto \mathbf{0},$$

$$(0, 0, 0, 0, 1, -1) \mapsto \mathbf{0}.$$

Na base

$$\begin{aligned} \mathcal{B} &= \{\mathbf{N}^2\mathbf{w}_1, \mathbf{N}\mathbf{w}_1, \mathbf{w}_1, \mathbf{N}\mathbf{w}_2, \mathbf{w}_2, \mathbf{w}_3\} \\ &= \{(-2, 0, -2, 0, -2, -2), (-2, 0, 0, 1, 1, 1), (0, 1, 0, 0, 0, 0) \\ &\quad (-1, 0, -1, 1, 0, 0), (0, 0, 1, 0, 0, 0), (0, 0, 0, 0, 1, -1)\}, \end{aligned}$$

temos a representação

$$[\mathbf{N}]_\delta = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

**3.4. Demonstrações alternativas.** Podemos fazer demonstrações alternativas de vários passos da demonstração de [Teorema 3.2](#) e até mesma da próxima demonstração como um todo. Isso é discutido nas seguintes observações.

OBSERVAÇÃO 3.4. . Uma demonstração de [\(3.3\)](#) se baseia na [Proposição 3.1](#), mas podemos dar uma demonstração mais direta. De fato, suponha que

$$\lambda_1 \llbracket \mathbf{N}\mathbf{w}_1 \rrbracket_{k-2} + \cdots + \lambda_{m_1} \llbracket \mathbf{N}\mathbf{w}_{m_1} \rrbracket_{k-2} = \llbracket \mathbf{0} \rrbracket_{k-2}.$$

Como  $\llbracket \mathbf{0} \rrbracket_{k-2} = \ker(\mathbf{N}^{k-2})$ , então

$$\mathbf{N}(\lambda_1 \mathbf{w}_1 + \cdots + \lambda_{m_1} \mathbf{w}_{m_1}) \in \ker(\mathbf{N}^{k-2}),$$

o que significa que

$$\lambda_1 \mathbf{w}_1 + \cdots + \lambda_{m_1} \mathbf{w}_{m_1} \in \ker(\mathbf{N}^{k-1}).$$

Mas, na classe de equivalência módulo  $\ker(\mathbf{N}^{k-1})$ , isso nos dá

$$\lambda_1 \llbracket \mathbf{w}_1 \rrbracket_{k-1} + \cdots + \lambda_{m_1} \llbracket \mathbf{w}_{m_1} \rrbracket_{k-1} = \llbracket \mathbf{0} \rrbracket_{k-1}.$$

Como  $\{\{\llbracket \mathbf{w}_1 \rrbracket_{k-1}, \dots, \llbracket \mathbf{w}_{m_1} \rrbracket_{k-1}\}\}$  é LI, visto que é uma base de  $\ker(\mathbf{N}^k)/\ker(\mathbf{N}^{k-1})$ , então

$$\lambda_1 = \cdots = \lambda_{m_1} = 0,$$

mostrando que  $\{\{\llbracket \mathbf{N}\mathbf{w}_1 \rrbracket_{k-2}, \dots, \llbracket \mathbf{N}\mathbf{w}_{m_1} \rrbracket_{k-2}\}\}$  é LI.

OBSERVAÇÃO 3.5. Uma demonstração de [\(3.4\)](#) se baseia na [Proposição 3.2](#), mas também podemos dar uma demonstração mais direta. Considere, então, uma combinação linear tal que

$$\begin{aligned} \lambda_1 \llbracket [\mathbf{N}^j \mathbf{w}_1] \rrbracket_{k-j-1} + \cdots + \lambda_{m_1} \llbracket [\mathbf{N}^j \mathbf{w}_{m_1}] \rrbracket_{k-j-1} + \cdots + \lambda_{m_{j-1}+1} \llbracket [\mathbf{N}\mathbf{w}_{m_{j-1}+1}] \rrbracket_{k-j-1} \\ + \cdots + \lambda_{m_j} \llbracket [\mathbf{N}\mathbf{w}_{m_j}] \rrbracket_{k-j-1} = \llbracket \mathbf{0} \rrbracket_{k-j-1}. \end{aligned}$$

Isso significa que

$$\begin{aligned} \lambda_1 \mathbf{N}^j \mathbf{w}_1 + \cdots + \lambda_{m_1} \mathbf{N}^j \mathbf{w}_{m_1} + \cdots + \lambda_{m_{j-1}+1} \mathbf{N}\mathbf{w}_{m_{j-1}+1} \\ + \cdots + \lambda_{m_j} \mathbf{N}\mathbf{w}_{m_j} \in \ker(\mathbf{N}^{k-j-1}). \end{aligned}$$

Ou seja,

$$\lambda_1 \mathbf{N}^{j-1} \mathbf{w}_1 + \cdots + \lambda_{m_1} \mathbf{N}^{j-1} \mathbf{w}_{m_1} + \cdots + \lambda_{m_{j-1}+1} \mathbf{w}_{m_{j-1}+1} + \cdots + \lambda_{m_j} \mathbf{w}_{m_j} \in \ker(\mathbf{N}^{k-j}).$$

Mas isso, por sua vez, significa que

$$\lambda_1 \llbracket \mathbf{N}^{j-1} \mathbf{w}_1 \rrbracket_{k-j} + \cdots + \lambda_{m_1} \llbracket \mathbf{N}^{j-1} \mathbf{w}_{m_1} \rrbracket_{k-j} + \cdots + \lambda_{m_{j-1}+1} \llbracket \mathbf{w}_{m_{j-1}+1} \rrbracket_{k-j} + \cdots + \lambda_{m_j} \llbracket \mathbf{w}_{m_j} \rrbracket_{k-j} = \llbracket \mathbf{0} \rrbracket_{k-j}.$$

Como esses elementos formam um conjunto LI, deduzimos que

$$\lambda_1 = \cdots = \lambda_{m_j} = 0,$$

mostrando que o conjunto (3.4) é linearmente independente.

OBSERVAÇÃO 3.6. Na demonstração do Teorema 3.2, deduzimos que a união, em  $j$ , dos conjuntos  $\mathcal{B}'_j$  é linearmente independente a partir da Proposição 3.2. Mas podemos dar uma demonstração mais direta. De fato, suponha que

$$\sum_{j=1, \dots, k} \sum_{\mathbf{w} \in \mathcal{B}'_j} \lambda_{\mathbf{w}}^j \mathbf{w} = \mathbf{0}.$$

para determinados escalares  $\lambda_{\mathbf{w}}^j$ . Primeiro, escrevemos que

$$\sum_{\mathbf{w} \in \mathcal{B}'_1} \lambda_{\mathbf{w}}^1 \mathbf{w} = - \sum_{j=2, \dots, k} \sum_{\mathbf{w} \in \mathcal{B}'_j} \lambda_{\mathbf{w}}^j \mathbf{w} \in \ker(\mathbf{N}^{k-1}),$$

enquanto que

$$\sum_{\mathbf{w} \in \mathcal{B}'_1} \lambda_{\mathbf{w}}^1 \mathbf{w} \in X \supset \ker(\mathbf{N}^{k-1}).$$

Então, na verdade,

$$\sum_{\mathbf{w} \in \mathcal{B}'_1} \lambda_{\mathbf{w}}^1 \mathbf{w} \in \ker(\mathbf{N}^{k-1}).$$

Isso significa, em termos da relação de equivalência módulo  $\ker(\mathbf{N}^{k-1})$ , que

$$\sum_{\mathbf{w} \in \mathcal{B}'_1} \lambda_{\mathbf{w}}^1 \llbracket \mathbf{w} \rrbracket_{k-1} = \llbracket \mathbf{0} \rrbracket_{k-1}.$$

Mas como  $\{\llbracket \mathbf{w} \rrbracket_{k-1}; \mathbf{w} \in \mathcal{B}'_1\}$  é base de  $\ker(\mathbf{N}^k)/\ker(\mathbf{N}^{k-1})$ , então  $\lambda_{\mathbf{w}}^1 = 0$ , para todo  $\mathbf{w} \in \mathcal{B}'_1$ .

Da mesma forma, a cada passo  $i = 1, \dots, k$ , escrevemos

$$\sum_{\mathbf{w} \in \mathcal{B}'_i} \lambda_{\mathbf{w}}^i \mathbf{w} = - \sum_{j=i+1, \dots, k} \sum_{\mathbf{w} \in \mathcal{B}'_j} \lambda_{\mathbf{w}}^j \mathbf{w} \in \ker(\mathbf{N}^{k-i}),$$

com

$$\sum_{\mathbf{w} \in \mathcal{B}'_i} \lambda_{\mathbf{w}}^i \mathbf{w} \in \ker(\mathbf{N}^{k-i+1}),$$

de modo que, na verdade,

$$\sum_{\mathbf{w} \in \mathcal{B}'_i} \lambda_{\mathbf{w}}^i [\mathbf{w}]_{k-i} = [\mathbf{0}]_{k-i}$$

e, por  $\{[\mathbf{w}]_{k-i}; \mathbf{w} \in \mathcal{B}'_i\}$  ser base de  $\ker(\mathbf{N}^{k-i+1})/\ker(\mathbf{N}^{k-i})$ , segue que  $\lambda_{\mathbf{w}}^i = 0$ , para todo  $\mathbf{w} \in \mathcal{B}'_i$ , concluindo, ao final, que todo o  $\mathcal{B}'$  é linearmente independente.

**OBSERVAÇÃO 3.7.** Uma outra demonstração do **Teorema 3.2**, equivalente à demonstração dada acima, mas sem fazer referência a espaço quociente pode ser dada pedindo que  $\{\mathbf{w}_1, \dots, \mathbf{w}_{m_1}\}$  gere um subespaço complementar ao subespaço  $\ker(\mathbf{N}^{k-1})$  em  $\ker(\mathbf{N}^k) = X$  e, em cada passo  $j = 1, \dots, k-1$ , mostrar que

$$\{\{N^j \mathbf{w}_1, \dots, N^j \mathbf{w}_{m_1}, \dots, N \mathbf{w}_{m_{j-1}+1}, \dots, N \mathbf{w}_{m_j}\}\}$$

é linearmente independente e gera um subespaço de  $\ker(\mathbf{N}^{k-j})$  que é transversal a  $\ker(\mathbf{N}^{k-j-1})$ , podendo, então, ser completado até um subespaço complementar a  $\ker(\mathbf{N}^{k-j-1})$  em  $\ker(\mathbf{N}^{k-j})$ , ou seja, tal que

$$\ker(\mathbf{N}^{k-j}) = \text{span} \{N^j \mathbf{w}_1, \dots, N^j \mathbf{w}_{m_1}, \dots, N \mathbf{w}_{m_{j-1}+1}, \dots, N \mathbf{w}_{m_j}, \mathbf{w}_{m_j+1}, \dots, \mathbf{w}_{m_{j+1}}\} \oplus \ker(\mathbf{N}^{k-j-1}).$$

Essa demonstração se baseia mais diretamente no resultado do **Proposição 3.2**.

**OBSERVAÇÃO 3.8.** Uma outra demonstração do **Teorema 3.2** pode ser dada via indução no índice de nilpotência. Nela, reduzimos um operador nilpotente  $\mathbf{N}$  de ordem  $k$ , em um espaço vetorial  $X$ , a um operador nilpotente de ordem  $k-1$  restrito à imagem  $\text{Im}(\mathbf{N})$  e assumindo que o resultado vale para operadores nilpotentes até ordem  $k-1$ . Completa-se esse conjunto LI até uma base de  $X$ . Como nada vem de graça, há, também, um certo trabalho em se mostrar esse último passo. Para essa demonstração, veja, por exemplo, [9, 1]. Veja, também, o **Exercício 5.17** e o **Exercício 5.18**, relacionados a essa demonstração.

## 4. Aplicações

### 4.1. O sistema massa-mola com amortecimento crítico.

## 5. Exercícios

## Exercícios

- 5.1.** Seja  $X$  um espaço vetorial de dimensão  $n \in \mathbb{N}$ . Mostre que, para qualquer inteiro  $0 \leq k < n$ , existe um operador linear  $\mathbf{N} : X \rightarrow X$  em  $X$  tal que  $\mathbf{N}^2 = 0$  e  $\dim(\text{Im}(\mathbf{N})) = k$ . Mostre, ainda, que não é possível ter um com  $k = n$ .
- 5.2.** Mostre que um operador linear  $\mathbf{N} : X \rightarrow X$  em um espaço vetorial  $X$  é 2-nilpotente, i.e.  $\mathbf{N}^2 = 0$ , se, e somente se,  $\text{Im}(\mathbf{N}) \subset \ker(\mathbf{N})$ .
- 5.3.** Seja  $\mathbf{N} : X \rightarrow X$  um operador linear em um espaço vetorial  $X$ . Suponha que  $\mathbf{N}$  seja nilpotente e que  $\ker(\mathbf{N}) \cap \text{Im}(\mathbf{N})$  seja trivial. Mostre que  $\mathbf{N}$  é o operador nulo,  $\mathbf{N} = 0$ .
- 5.4.** Seja  $\mathbf{N} \in \mathcal{L}(X)$  um operador nilpotente em um espaço vetorial  $X$ , com índice de nilpotência  $k \in \mathbb{N}$ . Mostre que  $\text{Im}(\mathbf{N}) \subset \ker(\mathbf{N}^{k-1})$ . Mais geralmente, mostre que  $\text{Im}(\mathbf{N}^j) \subset \ker(\mathbf{N}^{k-j})$ , para  $j = 0, 1, \dots, n$
- 5.5.** Em cada um dos itens a seguir, dê um exemplo de operador nilpotente  $\mathbf{N} \in \mathcal{L}(\mathbb{R}^n)$  de índice  $k$  sob as condições dadas, indicando exemplos de ciclos que caracterizam o operador.
- (1)  $n = 3, k = 3, \text{Im}(\mathbf{N}) = \ker(\mathbf{N}^2)$  e  $\text{Im}(\mathbf{N}^2) = \ker(\mathbf{N})$ ;
  - (2)  $n = 4, k = 2, \text{Im}(\mathbf{N}) = \ker(\mathbf{N})$ ;
  - (3)  $n = 4, k = 2, \text{Im}(\mathbf{N}) \subsetneq \ker(\mathbf{N})$ ;
  - (4)  $n = 4, k = 3, \text{Im}(\mathbf{N}) \subsetneq \ker(\mathbf{N}^2)$  e  $\text{Im}(\mathbf{N}^2) \subsetneq \ker(\mathbf{N})$ ;
  - (5)  $n = 5, k = 3, \text{Im}(\mathbf{N}) \subsetneq \ker(\mathbf{N}^2)$  e  $\text{Im}(\mathbf{N}^2) \subsetneq \ker(\mathbf{N})$ ;
  - (6)  $n = 6, k = 2, \text{Im}(\mathbf{N}) = \ker(\mathbf{N})$ ;
  - (7)  $n = 6, k = 2, \text{Im}(\mathbf{N}) \subsetneq \ker(\mathbf{N})$ ;
- 5.6.** É possível termos exemplos de operadores nilpotentes  $\mathbf{N} \in \mathcal{L}(\mathbb{R}^n)$  de índice  $k$  sob as seguintes condições? Caso positivo, dê um exemplo e exiba ciclos que caracterizam o operador. Em caso negativo, justifique.
- (1)  $n = 3, k = 3, \text{Im}(\mathbf{N}) \subsetneq \ker(\mathbf{N}^2)$  e  $\text{Im}(\mathbf{N}^2) \subsetneq \ker(\mathbf{N})$ ;
  - (2)  $n = 4, k = 3, \text{Im}(\mathbf{N}) = \ker(\mathbf{N}^2)$  e  $\text{Im}(\mathbf{N}^2) = \ker(\mathbf{N})$ ;
  - (3)  $n = 5, k = 3, \text{Im}(\mathbf{N}) = \ker(\mathbf{N}^2)$  e  $\text{Im}(\mathbf{N}^2) = \ker(\mathbf{N})$ ;
- 5.7.** Faça os detalhes da demonstração alternativa do **Teorema 3.2** descrita na **Observação 3.7**.

**5.8.** Considere o operador  $N : \mathbb{R}^4 \rightarrow \mathbb{R}^4$  associado à matriz

$$N = \begin{bmatrix} -123 & -37 & 185 & 40 \\ -3 & -1 & 5 & 1 \\ -80 & -24 & 120 & 26 \\ -12 & -4 & 20 & 4 \end{bmatrix}$$

Ache um ciclo ou conjunto de ciclos, a base  $\mathcal{B}$  associada a esse(s) ciclo(s) e a representação do operador nessa base, conforme o enunciado no **Teorema 3.2**.

**5.9.** Considere o operador  $N : \mathbb{R}^4 \rightarrow \mathbb{R}^4$  associado à matriz

$$N = \begin{bmatrix} 1 & 2 & 0 & 5 \\ 6 & -5 & 17 & -38 \\ -3 & 0 & -6 & 9 \\ -2 & 1 & -5 & 10 \end{bmatrix}$$

Ache um ciclo ou conjunto de ciclos, a base  $\mathcal{B}$  associada a esse(s) ciclo(s) e a representação do operador nessa base, conforme o enunciado no **Teorema 3.2**.

**5.10.** Considere o operador  $N : \mathbb{R}^6 \rightarrow \mathbb{R}^6$  associado à matriz

$$N = \begin{bmatrix} -3 & 1 & 3 & 0 & -1 & 0 \\ -2 & 1 & 2 & 0 & -1 & 0 \\ -1 & 0 & -1 & 1 & 1 & 0 \\ -5 & 1 & 7 & -1 & -3 & 1 \\ 2 & -1 & -6 & 2 & 3 & 0 \\ -1 & 0 & 1 & 0 & -1 & 1 \end{bmatrix}$$

Ache um ciclo ou conjunto de ciclos, a base  $\mathcal{B}$  associada a esse(s) ciclo(s) e a representação do operador nessa base, conforme o enunciado no **Teorema 3.2**.

**5.11.** Considere o operador  $N : \mathcal{P}_3 \rightarrow \mathcal{P}_3$  definido no espaço de polinômios de grau no máximo três por  $N(p) = p'' - p'$ , onde  $p'$  e  $p''$  indicam as derivadas de primeira e de segunda ordem de um polinômio  $p = p(x)$ . Mostre que  $N$  é nilpotente, com índice de nilpotência igual à dimensão do espaço e encontre um ciclo  $p_3 \mapsto p_2 \mapsto p_1 \mapsto p_0$ .

**5.12.** Verifique que qualquer matriz  $n \times n$  da forma a seguir é nilpotente com índice 2:

$$\begin{bmatrix} & a_1 & & a_1 & \cdots & & a_1 \\ & \vdots & & \vdots & \ddots & & \vdots \\ & & a_{n-1} & & a_{n-1} & \cdots & a_{n-1} \\ -a_1 - \cdots - a_{n-1} & -a_1 - \cdots - a_{n-1} & \cdots & -a_1 - \cdots - a_{n-1} & & & \end{bmatrix}.$$

- 5.13.** Verifique que qualquer matriz  $n \times n$  da forma a seguir é cíclica, ou seja, nilpotente com índice  $n$ :

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

- 5.14.** Considere o operador  $N : \mathcal{P}_3 \rightarrow \mathcal{P}_3$  definido no espaço de polinômios de grau no máximo três por  $N(p) = p''' - p''$ , onde  $p''$  e  $p'''$  indicam as derivadas de segunda e de terceira ordem de um polinômio  $p = p(x)$ . Encontre o índice de nilpotência de  $N$  e ciclos que, juntos, foram uma base de  $\mathcal{P}_3$ .
- 5.15.** Seja  $N : X \rightarrow X$  um operador nilpotente de índice  $k$ , em um espaço vetorial  $X$ . Mostre que  $I - N$  é invertível, com inversa

$$(I - N)^{-1} = I + N + N^2 + \cdots + N^{k-1} = \sum_{j=0}^{k-1} N^j.$$

Observe a semelhança com a expansão  $(1 - \theta)^{-1} = \sum_{j=0,1,\dots} \theta^j$ .

- 5.16.** Seja  $N : X \rightarrow X$  um operador nilpotente de índice  $k$ , em um espaço vetorial  $X$ . Mostre que  $I + N$  é invertível e encontre uma fórmula para a inversa  $(I + N)^{-1}$  em termos de potências de  $N$ .
- 5.17.** Seja  $N : X \rightarrow X$  um operador nilpotente de índice 2, em um espaço vetorial  $X$  de dimensão finita, i.e.  $N^2 = 0$ . Sejam  $m_1 = \dim(\text{Im}(N))$  e  $m_2 = \dim(\ker(N))$ . Considere uma base  $\{\{N\mathbf{w}_1, \dots, N\mathbf{w}_{m_1}\}\}$  da imagem  $\text{Im}(N)$  de  $N$  e uma base  $\{\{\mathbf{w}_{m_1+1}, \dots, \mathbf{w}_{m_2}\}\}$  do núcleo  $\ker(N)$  de  $N$ . Mostre que

$$\{\{N\mathbf{w}_1, \mathbf{w}_1, N\mathbf{w}_2, \mathbf{w}_2, \dots, N\mathbf{w}_{m_1}, \mathbf{w}_{m_1}, \mathbf{w}_{m_1+1}, \dots, \mathbf{w}_{m_2}\}\}$$

é uma base de  $X$ .

- 5.18.** Seja  $N : X \rightarrow X$  um operador nilpotente de índice 3, em um espaço vetorial  $X$  de dimensão finita, i.e.  $N^3 = 0$ . Mostre que  $N : \text{Im}(N) \rightarrow \text{Im}(N)$  é um operador nilpotente de índice 2 e, portanto,  $\text{Im}(N)$  possui uma base da forma

$$\{\{\mathbf{w}_1, \dots, \mathbf{w}_{m_1}, N\mathbf{w}_1, \dots, N\mathbf{w}_{m_1}, \mathbf{w}_{m_1+1}, \dots, \mathbf{w}_{m_2}\}\}$$

onde  $m_1 = \dim(\ker(\mathbf{N}^3)/\ker(\mathbf{N}^2))$  e  $m_2 = \dim(\ker(\mathbf{N}^2)/\ker(\mathbf{N}))$ . Escolhendo, agora, uma base  $\{\mathbf{w}_{m_2+1}, \dots, \mathbf{w}_{m_3}\}$  de  $\ker(\mathbf{N})$ , mostre que

$$\left\{ \left\{ \mathbf{N}^2 \mathbf{w}_1, \mathbf{N} \mathbf{w}_1, \mathbf{w}_1, \dots, \mathbf{N}^2 \mathbf{w}_{m_1}, \mathbf{N} \mathbf{w}_{m_1}, \mathbf{w}_{m_1}, \right. \right. \\ \left. \left. \mathbf{N} \mathbf{w}_{m_1+1}, \mathbf{w}_{m_1+1}, \dots, \mathbf{N} \mathbf{w}_{m_2}, \mathbf{w}_{m_2}, \mathbf{w}_{m_2+1}, \dots, \mathbf{w}_{m_3} \right\} \right\}$$

é uma base de  $X$ .

## CAPÍTULO 6

# Teoria Espectral e a Forma Normal de Jordan

### 1. Fundamentos

**1.1. Autovalores e autovetores.** Como visto nos primeiros cursos de Álgebra Linear, há certas “direções” privilegiadas no espaço, em que um operador age apenas esticando, contraindo e/ou refletindo os vetores naquela direção, ou seja, mudando a escala e/ou o sentido, mas sem alterar a direção. Além disso, vimos que pode haver várias dessas direções, com mudanças de escala e direção, diferentes ou não, e que, muitas vezes, caracterizam completamente o operador. Em alguns outros casos, envolvendo operadores nilpotentes, essa caracterização pode não ser completa, mas é um passo fundamental. Vamos, aqui, lembrar essa ideia, associada aos conceitos de autovalor, autovetor e autoespaço.

**DEFINIÇÃO 1.1.** *Seja  $T : X \rightarrow X$  um operador linear em um espaço vetorial  $X$  sobre um corpo  $\mathbb{K}$ . Um escalar  $\lambda \in \mathbb{K}$  é dito um **autovalor** de  $T$  quando existe um vetor  $\mathbf{u} \in X$  tal que  $\mathbf{u} \neq \mathbf{0}$  e*

$$T\mathbf{u} = \lambda\mathbf{u}.$$

*Um tal vetor não nulo  $\mathbf{u}$  é chamado de **autovetor** de  $T$  associado ao autovalor  $\lambda$ .*

**EXEMPLO 1.1.** Por exemplo, o operador  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  definido por

$$T(x, y) = (2x, -6y)$$

tem, como autovalores,  $\lambda_1 = 2$  e  $\lambda_2 = -6$ . Um autovetor de  $\lambda_1 = 2$  é o vetor

$$\mathbf{u}_1 = (1, 0),$$

ou qualquer múltiplo não nulo dele, ou seja, qualquer  $\mathbf{u} = (x, 0)$ , com  $x \neq 0$ , é um autovetor, satisfazendo

$$T(x, 0) = (2x, 0) = 2(x, 0).$$

Um autovetor de  $\lambda_2 = -6$  é o vetor

$$\mathbf{u}_2 = (0, 1),$$

ou qualquer  $\mathbf{u} = (0, y)$ , com  $y \neq 0$ , visto que

$$T(0, y) = (0, -6y) = -6(0, y).$$

Qualquer outro vetor não nulo muda de direção, visto que

$$T(x, y) = (2x, -6y),$$

cuja inclinação é  $-6y/2x = -3y/x$ , que é diferente de  $y/x$ , que a inclinação do vetor original.

**1.2. Autoespaços.** O conjunto de todos os autovetores associados a um determinado autovalor gera um subespaço linear chamado de autoespaço, conforme a definição a seguir.

DEFINIÇÃO 1.2. *Seja  $T : X \rightarrow X$  um operador linear em um espaço vetorial  $X$  sobre um corpo  $\mathbb{K}$ . Se  $\lambda \in \mathbb{K}$  é um autovalor de  $T$ , então o **autoespaço** de  $T$  associado a  $\lambda$  é o conjunto  $V(T, \lambda)$  definido por*

$$V(T, \lambda) = \ker(T - \lambda I) = \{\mathbf{u} \in X; T\mathbf{u} = \lambda\mathbf{u}\}.$$

OBSERVAÇÃO 1.1. Observe que, em  $V(T, \lambda)$ , na Definição 1.2, não se exige que  $\mathbf{u}$  seja diferente do vetor nulo. Ou seja, o autoespaço é formado pela união entre o conjunto de autovetores e o vetor nulo. Desta forma, ele se torna um subespaço. De fato, se  $\mathbf{u}, \mathbf{v} \in V(T, \lambda)$  e  $\alpha \in \mathbb{K}$ , então

$$T(\mathbf{u} + \alpha\mathbf{v}) = T\mathbf{u} + \alpha T\mathbf{v} = \lambda\mathbf{u} + \alpha\lambda\mathbf{v} = \lambda(\mathbf{u} + \alpha\mathbf{v}),$$

de modo que  $\mathbf{u} + \alpha\mathbf{v} \in V(T, \lambda)$ , mostrando que  $V(T, \lambda)$  é um subespaço.

OBSERVAÇÃO 1.2. No corpo dos reais, nem todo operador possui autovalor. O exemplo clássico é a rotação, digamos  $R : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  dado por

$$R(x, y) = (-y, x), \quad (x, y) \in \mathbb{R}^2,$$

que é a rotação de noventa graus no sentido trigonométrico. Se  $R\mathbf{u} = \lambda\mathbf{u}$ , para algum  $\mathbf{u} = (x, y)$ , então devemos ter  $(-y, x) = \lambda(x, y)$ . Nesse caso,  $-y = \lambda x$  e  $x = \lambda y$ , de modo que  $x = \lambda y = -\lambda(-y) = -\lambda(\lambda x) = -\lambda^2 x$ . Se  $x = 0$ , então  $y = -\lambda x = 0$  e  $\mathbf{u} = (0, 0)$  não pode ser autovetor. Se  $x \neq 0$ , então  $\lambda^2 = -1$ , que por sua vez, não tem solução real. Ou seja, não existe autovalor real  $\lambda$  de  $R$ , pois este deveria satisfazer  $\lambda^2 = -1$ . Na base canônica  $e = \{(1, 0), (0, 1)\}$ , temos a representação

$$[R]_e = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

OBSERVAÇÃO 1.3. Nos complexos, no entanto, o operador  $R : \mathbb{C}^2 \rightarrow \mathbb{C}^2$ , dado por

$$R(\xi, \eta) = (-\eta, \xi), \quad (\xi, \eta) \in \mathbb{C}^2,$$

que é a *complexificação* do operador rotação da **Observação 1.2**, tem autovalores complexos  $\pm i$ . Para achar o autoespaço, devemos resolver

$$\begin{cases} -\eta = \pm i\xi, \\ \xi = \pm i\eta, \end{cases}$$

Multiplicando a primeira equação por  $-i$ , obtemos a segunda equação, ou seja, nos complexos, as duas equações são redundantes. Isso define os autoespaços:

$$V(\mathbb{T}, i) = \{(\xi, \eta); \eta = -i\xi\} = \{(\xi, -i\xi); \xi \in \mathbb{C}\} = \text{span}\{(1, -i)\}$$

e

$$V(\mathbb{T}, -i) = \{(\xi, \eta); \eta = i\xi\} = \{(\xi, i\xi); \xi \in \mathbb{C}\} = \text{span}\{(1, i)\}$$

ambos de dimensão 1. Como a dimensão do espaço é dois, temos

$$\mathbb{C}^2 = V(\mathbb{T}, i) \oplus V(\mathbb{T}, -i)$$

Além disso, na base  $\mathfrak{b} = \{(1, -i), (1, i)\}$ , temos

$$[\mathbb{R}]_{\mathfrak{b}} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}.$$

O que faz isso funcionar é o fato de  $\mathbb{C}$  ser um corpo algebricamente fechado, o que significa dizer que qualquer polinômio não constante sobre os complexos possui raiz. Isso será explorado na próxima seção para garantir que todo operador linear sobre um espaço vetorial de dimensão finita possui autovalor.

Um resultado que é trivial mas que não custa ressaltar é o de que um autoespaço é um subespaço invariante pela transformação.

**TEOREMA 1.1.** *Seja  $\mathbb{T} \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  e suponha que  $\lambda$  seja um autovalor de  $\mathbb{T}$ . Então o autoespaço  $V(\mathbb{T}, \lambda) = \ker(\mathbb{T} - \lambda\mathbb{I})$  é um subespaço invariante por  $\mathbb{T}$ , com  $\mathbb{T}(V(\mathbb{T}, \lambda)) = V(\mathbb{T}, \lambda)$ , se  $\lambda \neq 0$ , e  $\mathbb{T}(V(\mathbb{T}, \lambda)) = \{\mathbf{0}\}$ , se  $\lambda = 0$ .*

**DEMONSTRAÇÃO.** De fato, se  $\mathbf{u} \in V(\mathbb{T}, \lambda)$ , então  $\mathbb{T}\mathbf{u} = \lambda\mathbf{u} \in V(\mathbb{T}, \lambda)$ , mostrando que  $V(\mathbb{T}, \lambda)$  é invariante. Caso  $\lambda = 0$ , então naturalmente  $\mathbb{T}\mathbf{u} = \mathbf{0}$ , para todo  $\mathbf{u} \in V(\mathbb{T}, \lambda)$ , de modo que  $\mathbb{T}(V(\mathbb{T}, \lambda)) = \{\mathbf{0}\}$ . Se  $\lambda \neq 0$ , dado  $\mathbf{u} \in V(\mathbb{T}, \lambda)$ , basta tomar  $\mathbf{v} = \mathbf{u}/\lambda$  para obter  $\mathbb{T}\mathbf{v} = \mathbb{T}\mathbf{u}/\lambda = \lambda\mathbf{u}/\lambda = \mathbf{u}$ , mostrando que  $\mathbb{T}(V(\mathbb{T}, \lambda)) = V(\mathbb{T}, \lambda)$ .  $\square$

**1.3. Espectro de um operador.** O conjunto de autovalores é de fundamental importância no estudo de um operador. Vamos ver que, a partir dele, pelo menos em dimensão finita, podemos caracterizar completamente um operador, obtendo que ele é similar a uma forma *canônica* dentro de uma classe finita bem definida de formas canônicas. Ou seja, conseguimos identificar por completo um conjunto de representantes do espaço quociente  $\mathcal{L}(X)/\sim$ , segundo a relação de similaridade de operadores.

Em dimensão finita, esse conjunto de autovalores leva o nome especial de *espectro*. Em dimensão infinita, no entanto, esse conjunto de autovalores não é suficiente. Por isso o conceito de espectro é generalizado de uma maneira ligeiramente diferente, definida como se segue.

**DEFINIÇÃO 1.3.** *Seja  $X$  um espaço vetorial sobre um corpo  $\mathbb{K}$  real ou complexo e seja  $T \in \mathcal{L}(X)$  um operador em  $X$ . O **espectro** de  $T$ , denotado por  $\sigma(T)$ , é definido por*

$$\sigma(T) = \{\lambda \in \mathbb{K}; (T - \lambda I) \text{ é um isomorfismo de } X\}.$$

Como dito acima, em dimensão finita, isso é equivalente ao conjunto de autovalores.

**TEOREMA 1.2.** *Seja  $X$  um espaço vetorial sobre um corpo  $\mathbb{K}$  real ou complexo e seja  $T \in \mathcal{L}(X)$  um operador em  $X$ . Supondo  $X$  de dimensão finita, o espectro  $\sigma(T)$  de  $T$  é formado pelo conjunto de autovalores de  $X$ .*

**DEMONSTRAÇÃO.** Vimos, na **Definição 1.1**, que um escalar  $\lambda$  é um autovalor de um operador  $T \in \mathcal{L}(X)$  quando existe um vetor não nulo  $\mathbf{u}$ , chamado autovetor, tal que

$$T\mathbf{u} = \lambda\mathbf{u}.$$

O conjunto de autovetores gera um subespaço chamado de autoespaço e que é caracterizado pelo núcleo

$$\ker(T - \lambda I).$$

Quando núcleo de  $T - \lambda I$  é não trivial,  $\lambda$  é um autovalor. Quando esse núcleo é trivial,  $\lambda$  não é um autovalor. Graças ao **Teorema 1.2**, podemos caracterizar um autovalor  $\lambda$  dependendo de  $T - \lambda I$  ser ou não uma bijeção. Como  $T - \lambda I$  é linear, isso é equivalente a ele ser um isomorfismo de espaços vetoriais. Portanto, o espectro é exatamente o conjunto de autovalores.  $\square$

**OBSERVAÇÃO 1.4.** Em dimensão infinita, podemos ter o núcleo de  $T - \lambda I$  não trivial, ou seja,  $T - \lambda I$  injetivo, mas sem que  $T - \lambda I$  seja bijetivo. Nesse caso, o espectro contém não apenas autovalores, mas outros escalares também. Em espaços

*normados* de dimensão infinita, um conceito que veremos no **Capítulo 7**, podemos diferenciar os casos em que  $T - \lambda I$  não é sobrejetivo e com imagem podendo ser densa ou não. Assim, temos o espectro dividido em três partes. O *espectro pontual*  $\sigma_p(T)$  feito dos escalares  $\lambda$  tais que  $T - \lambda I$  não é injetivo, ou seja, os autovalores. O *espectro contínuo*  $\sigma_c(T)$  em que  $T - \lambda I$  é injetivo mas não é bijetivo e a sua imagem é densa no espaço (i.e. o fecho da imagem é o espaço todo). O *espectro residual*  $\sigma_r(T)$  em que  $T - \lambda I$  é injetivo mas a sua imagem não é nem densa no espaço todo.

De qualquer forma, para todo  $\lambda$  fora do espectro, temos que  $T - \lambda I$  é um isomorfismo do espaço. O conceito de *raio espectral* nos dá uma limitação do espectro.

DEFINIÇÃO 1.4. *Seja  $X$  um espaço vetorial sobre um corpo  $\mathbb{K}$  real ou complexo e seja  $T \in \mathcal{L}(X)$  um operador em  $X$ . O **raio espectral** de  $T$  é definido por*

$$\rho(T) = \sup_{\lambda \in \sigma(T)} |\lambda|.$$

OBSERVAÇÃO 1.5. Em dimensão finita, esse raio espectral é sempre finito e é o máximo dos módulos dos autovetores. Veremos mais adiante que o número de autovalores é, no máximo, igual a dimensão do espaço, digamos  $\lambda_1, \dots, \lambda_k$ , com  $k \leq n = \dim(X)$ . Nesse caso,  $\rho(T) = \max\{|\lambda_1|, \dots, |\lambda_k|\}$ . Veremos, no **Capítulo 7**, que, mesmo em dimensão finita, se consideramos um operador  $T$  limitado (veja **Definição 1.10**), então  $\rho(T)$  também é finito.

OBSERVAÇÃO 1.6. O espectro aparece de forma fundamental em conjunto com o tratamento de operadores via fórmula de Cauchy, de integração complexa. Veja, por exemplo, **Observação 5.1**.

## 2. Diagonalização no caso de um conjunto completo de autovetores

Se for possível encontrar uma base formada por autovetores, então decompor o espaço nos autoespaços associados aos autovalores caracteriza completamente o operador através de simples operadores de mudança de escala, em cada autoespaço. Do ponto de vista de representação matricial, o operador toma a forma diagonal, nessa base.

**2.1. Autoespaços de autovalores distintos.** Um resultado importante, que iremos explorar, é o de que autoespaços de autovalores distintos são independentes. Isso será fundamental para a decomposição do espaço em subespaços invariantes em que um determinado operador atua de uma forma mais simples. Primeiramente, vamos ver o caso de dois autoespaços, apenas para efeitos didáticos, visto que a demonstração no caso geral pode ser feita de maneira independente.

TEOREMA 2.1. *Seja  $T \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  sobre um corpo  $\mathbb{K}$ . Sejam  $\lambda, \mu \in \mathbb{K}$  autovalores distintos de  $T$ . Então*

$$V(T, \lambda) \cap V(T, \mu) = \{\mathbf{0}\},$$

onde  $V(T, \lambda) = \ker(T - \lambda I)$  e  $V(T, \mu) = \ker(T - \mu I)$  são os autoespaços associados a  $\lambda$  e  $\mu$ .

DEMONSTRAÇÃO. Suponha que  $\mathbf{u} \in V(T, \lambda) \cap V(T, \mu)$ . Nesse caso, temos  $T\mathbf{u} = \lambda\mathbf{u}$  e  $T\mathbf{u} = \mu\mathbf{u}$ , de modo que

$$\lambda\mathbf{u} = \mu\mathbf{u}.$$

Como  $\lambda \neq \mu$ , então  $\mathbf{u} = \mathbf{0}$ . □

Agora, vamos ao caso de autoespaços de vários autovalores distintos.

TEOREMA 2.2. *Seja  $T \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  sobre um corpo  $\mathbb{K}$ . Sejam  $\lambda_1, \dots, \lambda_k \in \mathbb{K}$  autovalores distintos de  $T$ . Então os autoespaços  $V(T, \lambda_j) = \ker(T - \lambda_j I)$  são subespaços independentes entre si, de modo que vale a soma direta*

$$V(T, \lambda_1) \oplus \dots \oplus V(T, \lambda_k).$$

DEMONSTRAÇÃO. Graças ao Teorema 2.1, basta mostrar que, se  $\mathbf{u}_j \in V(T, \lambda_j)$ ,  $j = 1, \dots, k$ , com

$$\mathbf{u}_1 + \mathbf{u}_2 + \dots + \mathbf{u}_k = \mathbf{0}, \tag{2.1}$$

então cada  $\mathbf{u}_j = \mathbf{0}$  é nulo. Como de hábito, podemos mostrar isso por indução em  $k$ . Se  $k = 1$ , só temos um autoespaço e a condição (2.1) se reduz a  $\mathbf{u}_1 = \mathbf{0}$ , que nos diz diretamente que  $\mathbf{u}_1$  deve ser nulo.

Agora, supondo o resultado válido até  $k - 1$ , aplicamos  $(T - \lambda_1 I)$  à expressão (2.1), de modo que

$$\mathbf{0} + (T - \lambda_1 I)\mathbf{u}_2 + \dots + (T - \lambda_1 I)\mathbf{u}_k = \mathbf{0}.$$

Como cada  $\mathbf{u}_j$  é autovetor, isso nos dá

$$(\lambda_2 - \lambda_1)\mathbf{u}_2 + \dots + (\lambda_k - \lambda_1)\mathbf{u}_k = \mathbf{0}.$$

Pela hipótese de indução, isso significa que

$$(\lambda_j - \lambda_1)\mathbf{u}_j = \mathbf{0},$$

para  $j = 2, \dots, k$ . Como os autovalores são distintos, então

$$\mathbf{u}_j = \mathbf{0},$$

para  $j = 2, \dots, k$ . Por fim, como todos esses vetores se anulam, (2.1) nos diz que

$$\mathbf{u}_1 = \mathbf{0},$$

completando a demonstração.  $\square$

## 2.2. Diagonalização.

**TEOREMA 2.3.** *Seja  $T \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  de dimensão finita  $n = \dim(X) \in \mathbb{N}$ . Suponha que  $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  seja uma base formada por autovetores de  $T$ , associados a autovalores  $\lambda_1, \dots, \lambda_n$ , distintos ou não. Então, a representação  $[T]_{\mathcal{B}}$  de  $T$  nessa base é a matriz diagonal*

$$[T]_{\mathcal{B}} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} \quad (2.2)$$

**DEMONSTRAÇÃO.** Seja  $\mathcal{B}^* = \{\mathbf{f}_1, \dots, \mathbf{f}_n\}$  a base dual a  $\mathcal{B}$ . Isso significa que  $\langle \mathbf{f}_i, \mathbf{w}_j \rangle = \delta_{ij}$  é o Delta de Kronecker. Agora, a matriz  $[T]_{\mathcal{B}} = (a_{ij})_{i,j=1}^n$  é dada pelos coeficientes  $a_{ij} = \langle \mathbf{f}_i, T\mathbf{w}_j \rangle$ . Assim, usando que cada  $\mathbf{w}_j$  é um autovetor, temos

$$a_{ij} = \langle \mathbf{f}_i, T\mathbf{w}_j \rangle = \langle \mathbf{f}_i, \lambda_j \mathbf{w}_j \rangle = \lambda_j \langle \mathbf{f}_i, \mathbf{w}_j \rangle = \lambda_j \delta_{ij},$$

que é exatamente a matriz (2.2).  $\square$

No Teorema 2.3, definindo  $P_j$  como sendo o operador projeção sobre  $\text{span}\{\mathbf{w}_j\}$  ao longo de  $\text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_{j-1}, \mathbf{w}_{j+1}, \dots, \mathbf{w}_n\}$ , podemos escrever

$$I = P_1 + \cdots + P_n$$

e

$$T = \lambda_1 P_1 + \cdots + \lambda_n P_n.$$

Mas os autovalores podem ser repetidos, com autoespaços de dimensão maior do que um, sendo mais natural juntar os autovetores iguais nos seus respectivos autoespaços e reescrever isso da seguinte maneira.

**TEOREMA 2.4.** *Seja  $T \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  de dimensão finita  $n = \dim(X) \in \mathbb{N}$ , com autovalores  $\lambda_1, \dots, \lambda_k$ , com  $1 \leq k \leq n$ , com autoespaços de dimensão  $n_j = \dim(V(T, \lambda_j))$ ,  $j = 1, \dots, k$ . Suponha que*

$$n = n_1 + \cdots + n_k. \quad (2.3)$$

Então

$$X = V(T, \lambda_1) \oplus \cdots \oplus V(T, \lambda_k). \quad (2.4)$$

Além disso, definindo  $P_j \in \mathcal{L}(X)$  o operador projeção sobre  $V(T, \lambda_j)$  ao longo de  $\bigoplus_{i \neq j} V(T, \lambda_i)$ , então

$$I = P_1 + \cdots + P_k \quad (2.5)$$

e

$$\mathbf{T} = \lambda_1 \mathbf{P}_1 + \cdots + \lambda_k \mathbf{P}_k. \quad (2.6)$$

A representação matricial em uma base  $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_k$  formada pela união de bases arbitrárias  $\mathcal{B}_j = \{\mathbf{w}_1^j, \dots, \mathbf{w}_{n_j}^j\}$  de cada autoespaço  $V(\mathbf{T}, \lambda_j)$  toma a forma de blocos

$$[\mathbf{T}]_{\mathcal{B}} = \begin{bmatrix} \lambda_1 \mathbf{I}_{n_1} & & & \\ & \lambda_2 \mathbf{I}_{n_2} & & \\ & & \ddots & \\ & & & \lambda_k \mathbf{I}_{n_k} \end{bmatrix}. \quad (2.7)$$

DEMONSTRAÇÃO. Como os autoespaços são independentes, conforme visto no Teorema 2.2, e as dimensões deles somam até a dimensão do espaço, garantido pela condição (2.3), então o espaço todo  $X$  é soma direta dos autoespaços, provando (2.4). Graças a isso, as projeções  $\mathbf{P}_j$  somadas dão a identidade, nos dando (2.5).

Em cada autoespaço, temos  $\mathbf{T}\mathbf{u} = \lambda_j \mathbf{u}$ , para todo  $\mathbf{u} \in V(\mathbf{T}, \lambda_j)$ , o que nos dá, em geral, que

$$\mathbf{T}\mathbf{P}_j \mathbf{u} = \lambda_j \mathbf{P}_j \mathbf{u},$$

para todo  $\mathbf{u} \in X$ . Assim, usando (2.5), temos

$$\mathbf{T}\mathbf{u} = \mathbf{T}(\mathbf{P}_1 \mathbf{u} + \cdots + \mathbf{P}_k \mathbf{u}) = \lambda_1 \mathbf{P}_1 \mathbf{u} + \cdots + \lambda_k \mathbf{P}_k \mathbf{u} = (\lambda_1 \mathbf{P}_1 + \cdots + \lambda_k \mathbf{P}_k) \mathbf{u},$$

o que significa dizer que vale a identidade (2.6).

Por fim, na base  $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_k$  conforme descrita no enunciado, sendo  $\mathcal{B}^*$  a base dual com funcionais  $\mathbf{f}_i^m$ ,  $i = 1, \dots, n_m$ ,  $m = 1, \dots, k$ , temos

$$\langle \mathbf{f}_i^m, \mathbf{T}\mathbf{w}_l^j \rangle = \lambda_j \langle \mathbf{f}_i^m, \mathbf{w}_l^j \rangle = \lambda_j \delta_{mj} \delta_{il},$$

ou seja,  $\langle \mathbf{f}_i^m, \mathbf{T}\mathbf{w}_l^j \rangle = 0$ , em diferentes blocos  $m \neq j$ ;  $\langle \mathbf{f}_i^j, \mathbf{T}\mathbf{w}_l^j \rangle = 0$ , no mesmo bloco  $m = j$  mas fora da diagonal, i.e.  $i \neq l$ ; e apenas  $\langle \mathbf{f}_i^j, \mathbf{T}\mathbf{w}_i^j \rangle = \lambda_j$ , no mesmo bloco  $m = j$  e na diagonal  $l = i$ , provando (2.7).  $\square$

A condição de termos uma base formada por autovetores pode ser garantida caso tenhamos  $n$  autovalores distintos, onde  $n$  é a dimensão do espaço. Isso é consequência imediata do Teorema 2.4.

COROLÁRIO 2.1. *Suponha que  $\mathbf{T} \in \mathcal{L}(X)$  tenha  $n$  autovalores distintos  $\lambda_1, \dots, \lambda_n$ , associados a autovetores  $\mathbf{w}_1, \dots, \mathbf{w}_n$ , respectivamente, em um espaço vetorial  $X$  de dimensão finita  $n \in \mathbb{N}$ . Então  $\mathcal{B} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  forma uma base de  $X$  de autovetores de  $\mathbf{T}$ , de forma que o Teorema 2.3 se aplica.*

DEMONSTRAÇÃO. Cada autoespaço tem dimensão no mínimo um. Como são  $n$  autovalores, então necessariamente a dimensão  $n_j$  de cada um deles deve ser exatamente  $n_j = 1$ , com  $n_1 + \dots + n_n = n$ , de modo que podemos aplicar o Teorema 2.4.  $\square$

EXEMPLO 2.1. Considere o operador  $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  dado por

$$T(x, y, z) = T_{\mathbf{A}}(x, y, z) = (2x - y + z, x + z, x - y + 2z),$$

associado à matriz

$$[T]_e = \mathbf{A} = \begin{bmatrix} 2 & -1 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & 2 \end{bmatrix},$$

na base canônica  $e$ . Os autovalores são 1 e 2, com

$$V(T, 1) = \{y = x + z\} = \{(x, y, z); y = x + z\} = \{(r, r + s, s); r, s \in \mathbb{R}\}$$

e

$$V(T, 2) = \{z = y = x\} = \{(x, y, z); z = x = y\} = \{(s, s, s); s \in \mathbb{R}\}.$$

Observe que

$$n_1 = \dim(V(T, 1)) = 2, \quad n_2 = \dim(V(T, 2)) = 1, \quad n_1 + n_2 = 3 = n = \dim(X).$$

Portanto, a operador é diagonalizável. Podemos tomar  $\{(1, 1, 0), (0, 1, 1)\}$  como base de  $V(T, 1)$  e  $\{(1, 1, 1)\}$  como base de  $V(T, 2)$ . Assim, temos a base de autovetores

$$\mathcal{b} = \{(1, 1, 0), (0, 1, 1), (1, 1, 1)\}.$$

que gera subespaços invariantes. Para obtermos a representação de  $T$  na base  $\mathcal{b}$ , aplicamos o operador a cada vetor da base e representamos o resultado em termos dos vetores da base,

$$T(1, 1, 0) = (1, 1, 0) = 1(1, 1, 0) + 0(0, 1, 1) + 0(0, 1, 0),$$

$$T(0, 1, 1) = (0, 1, 1) = 0(1, 1, 0) + 1(0, 1, 1) + 0(0, 1, 0),$$

$$T(1, 1, 1) = (2, 2, 2) = 0(1, 1, 0) + 2(0, 1, 1) + 0(0, 1, 0).$$

Como era de se esperar, obtemos uma representação em forma diagonal, com os autovalores na diagonal, na ordem dos respectivos autovetores da base, i.e.

$$[T]_{\mathcal{b}} = [I]_{\mathcal{b}}^e [T]_e [I]_e^{\mathcal{b}} = \mathbf{M}^{-1} \mathbf{A} \mathbf{M} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix},$$

onde

$$\mathbf{M} = [I]_e^{\mathcal{b}} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

é a matriz mudança de base, da base  $\mathcal{b}$  pra base canônica  $e$ .

Vamos, agora, ver um exemplo com autovalores complexos.

EXEMPLO 2.2. Considere o operador  $T : \mathbb{C}^3 \rightarrow \mathbb{C}^3$  dado por

$$T(\xi, \eta, \zeta) = T_{\mathbf{A}}(\xi, \eta, \zeta) = (2\eta - \zeta, \xi + \eta - 2\zeta, -2\xi + 2\eta + \zeta),$$

associado à matriz

$$\mathbf{A} = [T]_e = \begin{bmatrix} 0 & 2 & -1 \\ 1 & 1 & -2 \\ -2 & 2 & 1 \end{bmatrix},$$

representando  $T$  na base canônica  $e$ . Os coeficientes reais e esse operador também pode ser visto como um operador linear restrito a  $\mathbb{R}^3$ , mas alguns dos seus autovalores,  $2, i, -i$ , são complexos, com

$$\begin{aligned} T(1, 1, 0) &= 2(1, 1, 0) \\ T(1 - 3i, 2 - i, 1 - 3i) &= i(1 - 3i, 2 - i, 1 - 3i) \\ T(1 + 3i, 2 + i, 1 + 3i) &= -i(1 + 3i, 2 + i, 1 + 3i). \end{aligned}$$

Escolhendo a base de autovetores

$$\mathcal{b} = \{(1, 1, 0), (1 - 3i, 2 - i, 1 - 3i), (1 + 3i, 2 + i, 1 + 3i)\},$$

obtemos a representação diagonal

$$[T]_{\mathcal{b}} = [I]_{\mathcal{b}}^e [T]_e [I]_e^{\mathcal{b}} = \mathbf{M}^{-1} \mathbf{A} \mathbf{M} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & i & 0 \\ 0 & 0 & -i \end{bmatrix},$$

onde  $\mathbf{M} = [I]_e^{\mathcal{b}}$  é a matriz mudança de base, da base  $\mathcal{b}$  para a base canônica  $e$ , dada por

$$\mathbf{M} = \begin{bmatrix} 1 & 1 - 3i & 1 + 3i \\ 1 & 2 - i & 2 + i \\ 0 & 1 - 3i & 1 + 3i \end{bmatrix}.$$

### 3. Autovalores complexos e forma triangular superior

Se existir uma base de autovetores, podemos, em princípio, obter uma representação na forma diagonal. Mas algumas questões ainda se impõem. Como podemos verificar a existência de uma base de autovetores e obter essa base, caso ela exista? O que podemos fazer caso uma base de autovetores não exista?

Nesse sentido, vamos verificar a existência de autovalores no corpo dos complexos e a existência de uma forma triangular. Em seguida vamos buscar refinar essa representação em uma forma diagonal em blocos, com blocos, de certa forma, mínimos.

Um ponto fundamental é que, no corpo dos complexos, todo operador linear em um espaço de dimensão finita possui pelo menos um autovalor. Isso sai do fato de que todo polinômio complexo possui pelo menos uma raiz. Vejamos esse resultado e outros que podem ser deduzidos dele sem muita dificuldade, como a existência de uma representação em forma triangular.

### 3.1. Existência de autovalores complexos.

**TEOREMA 3.1.** *Seja  $T : X \rightarrow X$  um operador linear em um espaço não trivial de dimensão finita  $X$  sobre o corpo  $\mathbb{C}$  dos complexos. Então  $T$  possui pelo menos um autovalor.*

**DEMONSTRAÇÃO.** Seja  $n = \dim(X)$ . Como  $X$  é de dimensão finita e é não trivial, então  $n \in \mathbb{N}$ . Seja  $\mathbf{u} \in X$  não nulo e considere a sequência de  $n + 1$  vetores  $\mathbf{u}, T\mathbf{u}, \dots, T^n\mathbf{u}$ . Como tem mais vetores do que a dimensão do espaço, eles formam, necessariamente, um conjunto linearmente dependente. Assim, existem escalares complexos  $c_0, \dots, c_n \in \mathbb{C}$ , nem todos nulos, tais que

$$c_0\mathbf{u} + c_1T\mathbf{u} + \dots + c_nT^n\mathbf{u} = \mathbf{0}.$$

Como  $\mathbf{u} \neq \mathbf{0}$ , então não podemos ter apenas  $c_0$  diferente de zero. Algum coeficiente  $c_k$ ,  $k \geq 1$ , deve ser diferente de zero. Seja  $k$  o maior coeficiente não nulo, de forma que  $1 \leq k \leq n$  com  $c_k \neq 0$  e

$$c_0\mathbf{u} + c_1T\mathbf{u} + \dots + c_kT^k\mathbf{u} = \mathbf{0}. \quad (3.1)$$

Considere, então, o polinômio de grau  $k$

$$p(z) = c_0 + c_1z + \dots + c_kz^k.$$

Como  $c_k \neq 0$  com  $k \geq 1$ , o polinômio  $p = p(z)$  não é constante. Como todo polinômio complexo não constante pode ser fatorado como produto de monômios, existem  $\zeta_1, \dots, \zeta_k \in \mathbb{C}$ , não necessariamente distintos, tais que

$$p(z) = c_k(z - \zeta_1)(z - \zeta_2) \cdots (z - \zeta_k).$$

Assim, podemos escrever

$$p(T) = c_k(T - \zeta_1 I)(T - \zeta_2 I) \cdots (T - \zeta_k I).$$

Da condição (3.1), temos

$$p(T)\mathbf{u} = c_k(T - \zeta_1 I)(T - \zeta_2 I) \cdots (T - \zeta_k I)\mathbf{u} = \mathbf{0}.$$

Nesse caso, devemos ter

$$(\mathbb{T} - \zeta_i \mathbb{I})\mathbf{v} = \mathbf{0},$$

para algum  $i = 1, \dots, n$  e algum  $\mathbf{v} \neq \mathbf{0}$ , onde

$$\mathbf{v} = \begin{cases} \mathbf{u}, & \text{se } i = k, \\ (\mathbb{T} - \zeta_{i+1}\mathbb{I}) \cdots (\mathbb{T} - \zeta_k\mathbb{I})\mathbf{u}, & \text{se } 1 \leq i < k. \end{cases}$$

Isso quer dizer que

$$\mathbb{T}\mathbf{v} = \zeta_i\mathbf{v},$$

com  $\mathbf{v} \neq \mathbf{0}$ , ou seja,  $\zeta_i$  é um autovalor com autovetor associado  $\mathbf{v}$ , provando que existe pelo menos um autovalor complexo.  $\square$

**OBSERVAÇÃO 3.1.** A demonstração do **Teorema 3.1** não quer dizer que todos as raízes  $\zeta_1, \dots, \zeta_n$  sejam autovalores nem que todos os autovalores são raízes do polinômio, como ocorre na fatoração do polinômio característico visto nos cursos iniciais de Álgebra Linear e que será visto mais pra frente. O polinômio construído acima depende do vetor inicial  $\mathbf{u}$ , não é uma propriedade única do operador. E isso também não quer dizer que um  $\mathbf{u}$  qualquer seja um autovetor. O autovetor  $\mathbf{v}$  da demonstração é obtido de  $\mathbf{u}$ , iterando os monômios obtidos a partir do próprio  $\mathbf{u}$ , como em uma “auto-correção” de rumo.

**OBSERVAÇÃO 3.2** (Subespaços de Krylov). Usamos a hipótese do espaço ser de dimensão finita, para deduzir que a sequência  $\mathbf{u}, \mathbb{T}\mathbf{u}, \dots, \mathbb{T}^n\mathbf{u}$  forma, necessariamente, um conjunto linearmente dependente. Essa sequência está associada aos chamados *subespaços de Krylov*, de ordem  $k$ , onde

$$\mathcal{K}_k(\mathbb{T}, \mathbf{u}) = \text{span}\{\mathbf{u}, \mathbb{T}\mathbf{u}, \dots, \mathbb{T}^{k-1}\mathbf{u}\}.$$

Eles formam um conjunto não decrescente de subespaços de  $X$ . E são encaixantes, no sentido de que  $\mathbb{T}\mathcal{K}_k(\mathbb{T}, \mathbf{u}) \subset \mathcal{K}_{k+1}(\mathbb{T}, \mathbf{u})$ . Subespaços de Krylov aparecem em métodos numéricos iterativos, para, por exemplo, encontrar autovalores e autovetores e na resolução de sistemas lineares de alta ordem, onde métodos diretos são mais custosos.

**OBSERVAÇÃO 3.3.** Apenas para efeito de curiosidade, vamos ver quem são o polinômio  $p$ , as raízes  $\zeta_i$  e o vetor  $\mathbf{v}$ , do **Teorema 3.1**, em um exemplo específico. Considere o operador  $\mathbb{T} : \mathbb{C}^3 \rightarrow \mathbb{C}^3$  definido por

$$\mathbb{T}(\xi, \eta, \zeta) = (2\xi, 3\eta, 4\zeta), \quad (\xi, \eta, \zeta) \in \mathbb{C}^3,$$

cujos autovalores são 2, 3 e 4. Seja  $\mathbf{u} = (1, 1, 0)$  e considere a sequência de vetores  $\mathbf{u}$ ,  $\mathbb{T}(\mathbf{u})$ ,  $\mathbb{T}^2(\mathbf{u})$ , e  $\mathbb{T}^3(\mathbf{u})$ , que nesse caso formam o conjunto

$$\{(1, 1, 0), (2, 3, 0), (4, 9, 0), (8, 18, 0)\}.$$

São quatro vetores em  $\mathbb{C}^3$ , portanto linearmente dependentes. (Podemos considerar isso em  $\mathbb{R}^3$ , também, mas, em geral, podemos não encontrar raízes reais.) Temos, por exemplo,

$$6(1, 1, 0) - 5(2, 3, 0) + (4, 9, 0) + 0(8, 18, 0) = (0, 0, 0).$$

Nesse caso, o polinômio  $p$  é

$$p(z) = 6 - 5z + z^2.$$

Suas raízes são reais, iguais a 2 e 3, que são parte dos autovalores, mas não são todos. Podemos escrever

$$p(z) = (z - 2)(z - 3) = (z - 3)(z - 2)$$

Como  $\mathbf{u} = (1, 1, 0)$  não é autovetor, então obtemos que

$$(\mathbf{T} - 3\mathbf{I})\mathbf{u} = (2, 3, 0) - 3(1, 1, 0) = (-1, 0, 0)$$

é autovetor associado ao autovalor 2 e

$$(\mathbf{T} - 2\mathbf{I})\mathbf{u} = (2, 3, 0) - 2(1, 1, 0) = (0, 1, 0)$$

é autovetor associado ao autovalor 3. Nesse exemplo, as duas raízes de  $p$  são autovalores do operador, que também possui outros autovalores. Observe que o vetor  $\mathbf{u}$  escolhido inicialmente pertence ao subespaço gerado pelos autoespaços de 2 e 3, por isso ele não “enxerga” o autovalor 4. A questão, agora, é se é possível obter um exemplo em que o polinômio possui raízes que não são autovalores do operador. Vejamos a próxima **Observação 3.4**.

**OBSERVAÇÃO 3.4.** Considere, agora, o operador nilpotente  $\mathbf{N}(x, y) = (y, 0)$ . Seja  $\mathbf{u} = (1, 0)$ . Então a sequência  $\mathbf{u}, \mathbf{N}\mathbf{u}, \mathbf{N}^2\mathbf{u}$  é  $(1, 0), (0, 0), (0, 0)$ . Podemos escolher  $c_0 = 0$  e  $c_1, c_2$  arbitrários que vale

$$c_0(1, 0) + c_1(0, 0) + c_2(0, 0) = (0, 0).$$

Assim,

$$p(z) = c_1z + c_2z^2 = c_2z \left( z + \frac{c_1}{c_2} \right) = c_2(z - \zeta_1)(z - \zeta_2),$$

com

$$\zeta_1 = 0, \quad \zeta_2 = -\frac{c_1}{c_2}.$$

Assim, encontraremos o autovalor nulo  $\zeta_1 = 0$  e uma outra raiz qualquer  $\zeta_2$ , visto que  $c_1$  e  $c_2$  são arbitrários. Escolhendo  $c_1 = 0$  e  $c_2 \neq 0$ , então  $\zeta_2 = 0$  e  $p(z) = c_2z^2$ . Mas para  $c_1, c_2 \neq 0$ , temos  $\zeta_2$  que não é autovalor.

**3.2. Forma triangular superior.** Obter a forma diagonal de um operador diagonalizável ou a forma de Jordan de um operador qualquer, em dimensão finita, não é algo tão imediato. Mas a forma triangular não requer muito mais esforço.

No caso de uma matriz triangular superior  $n \times n$ , ou seja, da forma

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

vemos que um vetor com as últimas colunas nulas é levado em outro vetor com (pelo menos) as mesmas últimas colunas nulas, i.e.

$$(x_1, x_2, \dots, x_j, 0, \dots, 0) \mapsto \left( \sum_{1 \leq k \leq j} a_{1k} x_k, \sum_{2 \leq k \leq j} a_{2k} x_k, \dots, a_{jj} x_j, 0, \dots, 0 \right).$$

Podemos ver isso, mais facilmente, usando uma estrutura de blocos,

$$\begin{bmatrix} \star & \star \\ \mathbf{0} & \star \end{bmatrix} \begin{pmatrix} \star \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \star \\ \mathbf{0} \end{pmatrix},$$

onde  $\mathbf{0}$  da matriz é um bloco  $(n-j) \times j$  e  $\mathbf{0}$  dos vetores são blocos  $1 \times j$ .

Em termos de subespaços vetoriais, podemos dizer que

$$\mathbb{T} \text{span}\{\mathbf{e}_1, \dots, \mathbf{e}_k\} \subset \text{span}\{\mathbf{e}_1, \dots, \mathbf{e}_k\},$$

ou, de outra forma, que  $\text{span}\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$  é *invariante* por  $\mathbb{T}$ , conforme visto para blocos na [Seção 5.4](#). Mais geralmente, temos a seguinte versão mais detalhada do [Teorema 5.3](#).

**TEOREMA 3.2.** *Seja  $\mathbb{T} : X \rightarrow X$  um operador linear em um espaço vetorial  $X$  de dimensão finita  $n = \dim(X) \in \mathbb{N}$ . Se  $\mathcal{b} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  é uma base de  $X$  tal que*

$$\mathbb{T}(\text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_j\}) \subset \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_j\},$$

*para todo  $j = 1, \dots, n$ , então a representação de  $\mathbb{T}$  na base  $\mathcal{b}$  tem a forma triangular*

$$[\mathbb{T}]_{\mathcal{b}} = \begin{bmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & \cdots & a_{2n} \\ \vdots & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & a_{nn} \end{bmatrix} \quad (3.2)$$

DEMONSTRAÇÃO. Basta observar que, dado um vetor

$$\mathbf{u} = x_1 \mathbf{w}_1 + \cdots + x_k \mathbf{w}_j \in \text{span}\{\mathbf{w}_1, \cdots, \mathbf{w}_j\}$$

a sua imagem  $\mathbb{T}\mathbf{u}$  também está em neste espaço gerado, i.e.

$$\mathbb{T}\mathbf{u} = y_1 \mathbf{w}_1 + \cdots + y_j \mathbf{w}_j \in \text{span}\{\mathbf{w}_1, \cdots, \mathbf{w}_j\},$$

para coeficientes apropriados, de forma que  $y_i = 0$ , para  $i > j$ , nos dando que os coeficientes  $a_{ij}$  de  $[\mathbb{T}]_{\mathfrak{b}}$  da parte inferior à diagonal,  $i > j$ , se anulam, restando uma forma triangular superior.  $\square$

OBSERVAÇÃO 3.5. Na forma triangular (3.2) do Teorema 3.2, os coeficientes de  $[\mathbb{T}]_{\mathfrak{b}}$  podem ser escritos como na Definição 3.5, usando a base dual  $\mathfrak{b}^* = \{\mathbf{f}_1, \dots, \mathbf{f}_n\}$  de  $\mathfrak{b}$ , apenas observando que  $\langle \mathbf{f}_i, \mathbf{w}_j \rangle = 0$  para  $i > j$ .

**3.3. Existência da forma triangular superior.** No caso de espaços vetoriais complexos, como a existência de autovalor está garantida, podemos usar um método indutivo para deduzir que um operador sempre possui uma forma triangular.

TEOREMA 3.3. *Seja  $\mathbb{T} : X \rightarrow X$  um espaço vetorial complexo  $X$  de dimensão finita  $n = \dim(X) \in \mathbb{N}$ . Então existe uma base  $\mathfrak{b}$  satisfazendo as condições do Teorema 3.2 e, portanto,  $\mathbb{T}$  possui a forma triangular superior nessa base.*

DEMONSTRAÇÃO. Se  $\dim(X) = 1$ , então  $[\mathbb{T}]_{\mathfrak{b}} = [a]$  para algum escalar  $a$ , em qualquer base, que é trivialmente triangular (i.e.  $\mathbb{T}(\text{span}\{\mathbf{w}_1\}) \subset \text{span}\{\mathbf{w}_1\}$ , para qualquer base  $\{\mathbf{w}_1\}$ ), sendo o resultado válido nesse caso.

Considere, agora,  $n = \dim(X) > 1$  e suponha o resultado válido em qualquer espaço de dimensão menor do que  $n$ . Vamos mostrar que também vale para  $n$ .

Seja, então,  $\mathbb{T} \in \mathcal{L}(X)$  arbitrário em  $X$ , onde  $\dim(X) = n$ . Como o espaço é complexo de dimensão finita, existe pelo menos um autovalor  $\lambda_1$  de  $\mathbb{T}$ . Nesse caso,  $(\mathbb{T} - \lambda_1 I)$  é um operador linear não injetivo, visto que os autovetores associados ao autovalor  $\lambda_1$  existem e estão no núcleo desse operador. Em outras palavras,

$$n_1 = \dim(\text{Im}(\mathbb{T} - \lambda_1 I)) = n - \dim(\ker(\mathbb{T} - \lambda_1 I)) < n.$$

Não custa ressaltar que, como toda imagem de um operador linear, a imagem  $\text{Im}(\mathbb{T} - \lambda_1 I)$  é um subespaço vetorial de  $X$ .

Além disso, o espaço  $\text{Im}(\mathbb{T} - \lambda_1 I)$  é invariante por  $\mathbb{T}$ . De fato, se  $\mathbf{u} = (\mathbb{T} - \lambda_1 I)\mathbf{v}$  para algum  $\mathbf{v} \in X$ , então

$$\mathbb{T}\mathbf{u} = \mathbb{T}(\mathbb{T} - \lambda_1 I)\mathbf{v} = \mathbb{T}^2 - \lambda_1 \mathbb{T}\mathbf{v} = (\mathbb{T} - \lambda_1 I)\mathbb{T}\mathbf{v}.$$

Ou seja,  $\mathbb{T}\mathbf{u}$  é a imagem de  $\mathbb{T}\mathbf{v}$  por  $(\mathbb{T} - \lambda_1 I)$ , o que significa dizer que

$$\mathbb{T}\mathbf{u} \in \text{Im}(\mathbb{T} - \lambda_1 I),$$

Portanto,  $\text{Im}(\mathbb{T} - \lambda_1 I)$  é invariante por  $\mathbb{T}$ .

Considere, agora,  $X_1 = \text{Im}(\mathbb{T} - \lambda_1 I)$  e  $\mathbb{T}_1 : X_1 \rightarrow X_1$  definido por

$$\mathbb{T}_1 \mathbf{u} = \mathbb{T} \mathbf{u}, \quad \forall \mathbf{u} \in X_1,$$

que, nada mais é, do que a restrição de  $\mathbb{T}$  ao subespaço vetorial  $X_1$ . Como  $\dim(X_1) = n_1 < n$ , temos que o resultado de existência de forma triangular vale em  $X_1$ . Em particular, vale para  $\mathbb{T}_1$  recém-definido. Seja, então,  $\mathcal{b}_1 = \{\mathbf{w}_1, \dots, \mathbf{w}_{n_1}\}$  uma base de  $X_1$  tal que

$$\mathbb{T}_1 \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_j\} \subset \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_j\}, \quad (3.3)$$

para qualquer  $j = 1, \dots, n_1$ . Observe que, como  $X_1$  é um subespaço de  $X$ , então os vetores  $\mathbf{w}_1, \dots, \mathbf{w}_{n_1}$  são vetores em  $X$ . Em seguida, estendemos essa base de  $X_1$  a uma base  $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  de  $X$ . A propriedade (3.3) significa o mesmo para  $\mathbb{T}$ , para  $1 \leq j \leq n_1$ . Falta o caso  $n_1 < j \leq n$ . Para isso, precisamos ver o que acontece ao aplicamos  $\mathbb{T}$  a cada novo vetor  $\mathbf{w}_j$ . Pare eles, temos

$$\mathbb{T} \mathbf{w}_j = \mathbb{T} \mathbf{w}_j - \lambda_1 \mathbf{w}_j + \lambda_1 \mathbf{w}_j = (\mathbb{T} - \lambda_1 I) \mathbf{w}_j + \lambda_1 \mathbf{w}_j.$$

Como

$$(\mathbb{T} - \lambda_1 I) \mathbf{w}_j \in \text{Im}(\mathbb{T} - \lambda_1 I) = X_1 = \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_{n_1}\},$$

segue que

$$\mathbb{T} \mathbf{w}_j \in \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_{n_1}\} + \text{span}\{\mathbf{w}_j\} \subset \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_{n_1}, \dots, \mathbf{w}_j\}.$$

Portanto, qualquer combinação linear dos vetores  $\mathbf{w}_i$  com  $i = 1, \dots, j$  é levada, por  $\mathbb{T}$ , a um elemento do espaço gerado por esses mesmos vetores, i.e.

$$\mathbb{T} \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_j\} \subset \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_j\},$$

agora para todo  $j = 1, \dots, n$ , completando a demonstração.  $\square$

EXEMPLO 3.1. Considere, novamente, o operador  $\mathbb{T} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ , do Exemplo 2.1, dado por

$$\mathbb{T}_{\mathbf{A}}(x, y, z) = (2x - y + z, x + z, x - y + 2z),$$

associado à matriz

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

Podemos verificar que a base

$$\mathcal{b} = \{(1, 1, 0), (0, 1, 0), (0, 0, 1)\}$$

gera subespaços invariantes encaixantes,

$$S_1 = \text{span}\{(1, 1, 0)\}, \quad S_2 = \text{span}\{(1, 1, 0), (0, 0, 1)\},$$

$$S_3 = \{(1, 1, 0), (0, 0, 1), (0, 1, 0)\} = \mathbb{R}^3.$$

Para obtermos a representação de  $\mathbb{T}$  na base  $\mathcal{b}$ , aplicamos o operador a cada vetor da base e representamos o resultado em termos dos vetores da base,

$$\begin{aligned} \mathbb{T}(1, 1, 0) &= (1, 1, 0) = 1(1, 1, 0) + 0(0, 0, 1) + 0(0, 1, 0), \\ \mathbb{T}(0, 0, 1) &= (1, 1, 2) = 1(1, 1, 0) + 2(0, 0, 1) + 0(0, 1, 0), \\ \mathbb{T}(0, 1, 0) &= (-1, 0, -1) = -1(1, 1, 0) - 1(0, 0, 1) + 1(0, 1, 0). \end{aligned}$$

Assim, obtemos a forma triangular superior

$$[\mathbb{T}]_{\mathcal{b}} = \mathbf{M}^{-1} \mathbf{A} \mathbf{M} = \begin{bmatrix} 1 & 1 & -1 \\ 0 & 2 & -1 \\ 0 & 0 & 1 \end{bmatrix},$$

onde

$$\mathbf{M} = [1]_{\mathcal{e}}^{\mathcal{b}} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

é a matriz mudança de base, da base  $\mathcal{b}$  pra base canônica  $\mathcal{e}$ .

**3.4. Autovalores e a diagonal da forma triangular.** De posse da forma triangular, podemos deduzir que todos os elementos da diagonal são autovalores, nos dando, portanto,  $n$  autovalores, contando multiplicidades, onde  $n$  é a dimensão do espaço. Os elementos da diagonal podem ser obtidos via base dual, conforme feito no [Definição 3.5](#).

**TEOREMA 3.4.** *Seja  $\mathbb{T} : X \rightarrow X$  um espaço vetorial complexo  $X$  de dimensão finita  $n = \dim(X) \in \mathbb{N}$  e seja  $\mathcal{b} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  uma base satisfazendo as condições do [Teorema 3.2](#), com  $\mathbb{T}$  possuindo forma triangular superior nessa base. Então, todos os elementos da diagonal dessa representação são autovalores de  $\mathbb{T}$ .*

**DEMONSTRAÇÃO.** Seja  $[\mathbb{T}]_{\mathcal{b}} = (a_{ij})_{ij}$  a forma triangular superior de  $\mathbb{T}$  nessa base, de modo que  $a_{ij} = 0$  quando  $i > j$ . De maneira geral, temos, por conta da forma triangular superior, que

$$\mathbb{T}\mathbf{w}_j = a_{1j}\mathbf{w}_1 + \dots + a_{nj}\mathbf{w}_n = a_{1j}\mathbf{w}_1 + \dots + a_{jj}\mathbf{w}_j, \quad (3.4)$$

Em particular, para  $j = 1$ , temos

$$\mathbb{T}\mathbf{w}_1 = a_{11}\mathbf{w}_1,$$

ou seja,  $\lambda_1 = a_{11}$  é autovalor, com autovetor  $\mathbf{w}_1$ .

Agora, para  $1 < j \leq n$ , observe, de (3.4) que, passando o último termo para o lado esquerdo,

$$(\mathbb{T} - a_{jj}I)\mathbf{w}_j = a_{1j}\mathbf{w}_1 + \cdots + a_{(j-1)j}\mathbf{w}_{j-1} \in \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_{j-1}\},$$

Além disso, para  $1 \leq i \leq j-1$ , temos, por conta da forma triangular superior, que

$$(\mathbb{T} - a_{jj}I)\mathbf{w}_i = a_{1i}\mathbf{w}_1 + \cdots + a_{ii}\mathbf{w}_i - a_{jj}\mathbf{w}_i \in \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_i\} \subset \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_{j-1}\}.$$

Logo,

$$(\mathbb{T} - a_{jj}I)\text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_j\} \subset \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_{j-1}\} \subsetneq \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_j\}.$$

Aplicando o Teorema 2.1 do Núcleo e da Imagem ao operador  $(\mathbb{T} - a_{jj}I)$  restrito ao subespaço  $j$ -dimensional  $\text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_j\}$ , vemos que

$$\begin{aligned} \dim(\ker(\mathbb{T} - a_{jj}I)) &= j - \dim(\text{Im}(\mathbb{T} - a_{jj}I)) \\ &\geq j - \dim(\text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_{j-1}\}) = j - (j-1) = 1. \end{aligned}$$

Portanto, existe pelo menos um vetor não nulo  $\mathbf{u}_j$  no núcleo de  $\mathbb{T} - a_{jj}I$ , o que significa que

$$(\mathbb{T} - a_{jj}I)\mathbf{u}_j = \mathbf{0},$$

que por sua vez nos dá

$$\mathbb{T}\mathbf{u}_j = a_{jj}\mathbf{u}_j,$$

com  $\mathbf{u}_j \neq \mathbf{0}$ . Em outras palavras,  $\lambda_j = a_{jj}$  é um autovalor, com autovetor associado  $\mathbf{u}_j \in \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_j\}$ .

Com isso vale para todo  $j = 1, \dots, n$ , a demonstração está completa.  $\square$

**OBSERVAÇÃO 3.6.** Lembremos que, usando a base  $\mathcal{b} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  e a base dual  $\mathcal{b}^* = \{\mathbf{f}_1, \dots, \mathbf{f}_n\}$ , no Teorema 3.4, os elementos  $a_{jj}$  da diagonal, que são autovalores, são dados por  $a_{jj} = \langle \mathbf{f}_j, \mathbf{w}_j \rangle$ .

O resultado acima diz que todos os valores da diagonal são autovalores, mas não quer dizer que todos os autovalores estão na diagonal. Ou seja, em princípio, poderia ter algum outro autovalor que não aparece em uma determinada forma triangular. É claro que, dado um autovalor, podemos começar a construir a base com um autovetor associado a esse autovalor, de forma que ele apareceria como primeiro elemento da diagonal dessa base assim construída. Mas isso não quer dizer, ainda, que esse processo abarcaria todos os autovalores. O resultado a seguir garante isso.

TEOREMA 3.5. *Seja  $\mathbb{T} : X \rightarrow X$  um espaço vetorial complexo  $X$  de dimensão finita  $n = \dim(X) \in \mathbb{N}$  e seja  $\lambda$  um autovalor de  $\mathbb{T}$ . Suponha que  $\mathcal{B} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  seja uma base satisfazendo as condições do Teorema 3.2, com  $\mathbb{T}$  possuindo forma triangular superior nessa base. Então, para algum inteiro  $1 \leq j \leq n$ , temos  $\lambda = a_{jj}$ , onde  $a_{jj}$  é um elemento da diagonal da representação  $[\mathbb{T}]_{\mathcal{B}} = (a_{ij})_{i,j=1}^{i,j=n}$  de  $\mathbb{T}$  nessa base. Ou seja, todos os autovalores de  $\mathbb{T}$  aparecem na diagonal de qualquer forma triangular superior de  $\mathbb{T}$ .*

DEMONSTRAÇÃO. A ideia é começar escrevendo um autovetor  $\mathbf{u}$ , de um autovalor  $\lambda$ , como

$$\mathbf{u} = x_1 \mathbf{w}_1 + \dots + x_j \mathbf{w}_j,$$

onde  $j$  é o maior inteiro tal que  $x_j$  é não nulo na representação de  $\mathbf{u}$  na base  $\mathcal{B}$ . Em seguida, observamos que, por um lado, como  $\mathbf{u}$  é autovetor,

$$\mathbb{T}\mathbf{u} = \lambda\mathbf{u} = \lambda x_1 \mathbf{w}_1 + \dots + \lambda x_j \mathbf{w}_j,$$

enquanto que, usando a forma triangular superior, temos

$$\begin{aligned} \mathbb{T}\mathbf{u} &= x_1 \mathbb{T}\mathbf{w}_1 + \dots + x_j \mathbb{T}\mathbf{w}_j \\ &= (a_{11}x_1 + a_{12}x_2 + \dots + a_{1j}x_j)\mathbf{w}_1 + (a_{22}x_2 + \dots + a_{2j}x_j)\mathbf{w}_2 + \dots + a_{jj}x_j\mathbf{w}_j. \end{aligned}$$

Como os coeficientes são únicos, temos

$$\lambda x_j = a_{jj}x_j,$$

sendo que  $x_j \neq 0$ , ou seja,

$$\lambda = a_{jj}.$$

□

OBSERVAÇÃO 3.7. Os resultados acima garantem que todos os elementos da diagonal de uma forma triangular são autovalores e todos os autovalores estão na diagonal de qualquer forma triangular. Mas não garantem, que, caso a diagonal tenha autovalores repetidos, as multiplicidades deles serão sempre as mesmas em qualquer forma triangular. Mais precisamente, se os autovalores são 2 e 3 em um espaço de dimensão três, não há nada garantindo, até o momento, que eu não possa ter uma diagonal com elementos 2, 2, 3, em uma base, e 2, 3, 3, em outra. Mas esse resultado é verdade, ou seja, todas as formas triangulares vão ter as mesmas diagonais, a menos de permutações. Em particular, se o autoespaço do autovalor 2 tem dimensão dois e o do autovalor 3 tem dimensão um, então todas as formas triangulares superiores terão, na sua diagonal, dois elementos iguais a 2 e apenas um elemento igual a 1. Não provaremos isso diretamente. Vamos ver isso como consequência do Teorema 4.6.

#### 4. A forma canônica de Jordan

**4.1. Autoespaços generalizados.** No caso de um operador nilpotente  $\mathbf{N} \in \mathcal{L}(X)$ , com índice de nilpotência  $k > 0$ , o único autovalor é  $\lambda = 0$ , mas o conjunto de autovetores não gera o espaço todo. De fato, vejamos o seguinte exemplo.

EXEMPLO 4.1. Seja  $\mathbf{N} \in \mathcal{L}(\mathbb{R}^2)$  dado por  $\mathbf{N}(x, y) = (y, x)$ . Então um par autovalor-autovetor  $(\lambda, \mathbf{u})$  satisfaz a equação vetorial  $\mathbf{N}\mathbf{u} = \lambda\mathbf{u}$ , que tem a forma

$$\mathbf{N}(x, y) = (y, 0) = \lambda(x, y).$$

Isso pode ser escrito na forma de sistema,

$$\begin{cases} y = \lambda x, \\ 0 = \lambda y, \end{cases}$$

Se  $\lambda \neq 0$ , então  $y = 0$  e, com isso,  $x = 0$  também, logo  $\lambda$  não pode ser autovalor. Se  $\lambda = 0$ , então  $y = 0$  e  $x$  pode ser qualquer real. Como um autovetor deve ser não nulo, então devemos ter  $x \neq 0$ ,  $y = 0$  e  $\lambda = 0$ . Ou seja, o único autovalor é  $\lambda = 0$  e o seu autoespaço é o eixo  $x$ ,

$$\begin{aligned} V(\mathbf{T}, 0) = \ker(\mathbf{T} - 0\mathbf{I}) &= \ker(\mathbf{T}) = \{(x, y) \in \mathbb{R}^2; y = 0\} = \{(x, 0) \in \mathbb{R}^2, x \in \mathbb{R}\} \\ &= \{\text{eixo } x\} = \text{span}\{(1, 0)\}. \end{aligned}$$

Mas sabemos que  $\mathbf{N}^2(0, 1) = (0, 0)$ . Na base canônica  $\mathbf{e} = \{(1, 0), (0, 1)\}$ , o operador toma a forma

$$[\mathbf{T}]_{\mathbf{e}} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

No exemplo acima, completamos a base com um vetor que não é autovetor, mas que de certa forma está associado ao autovalor  $\lambda = 0$ . Ele não está no núcleo de  $\mathbf{T} - \lambda\mathbf{I}$ , mas está no núcleo de alguma potência desse operador. Nesse sentido, fazemos a seguinte definição.

DEFINIÇÃO 4.1. Seja  $\mathbf{T} \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  sobre um corpo  $\mathbb{K}$ . Seja  $\lambda \in \mathbb{K}$  um autovalor de  $\mathbf{T}$ . Dizemos que  $\mathbf{u} \in X$  é um **autovetor generalizado** de índice  $k \in \mathbb{N}$  quando  $\mathbf{u} \in \ker(\mathbf{T} - \lambda\mathbf{I})^k \setminus \ker(\mathbf{T} - \lambda\mathbf{I})^{k-1}$ . O conjunto  $G(\mathbf{T}, \lambda)$  definido por

$$G(\mathbf{T}, \lambda) = \bigcup_{k \in \mathbb{N}} \ker(\mathbf{T} - \lambda\mathbf{I})^k$$

é chamado de **autoespaço generalizado** de  $\mathbf{T}$ , associado ao autovalor  $\lambda$ .

OBSERVAÇÃO 4.1. Naturalmente, no caso  $k = 1$ , temos  $\ker(\mathbb{T} - \lambda I)^k = \ker(\mathbb{T} - \lambda I) = V(\mathbb{T}, \lambda)$ , com  $\ker(\mathbb{T} - \lambda I)^{k-1} = \ker(I) = \{\mathbf{0}\}$ , de modo que um autovalor generalizado de índice  $k = 1$  é simplesmente um autovalor.

OBSERVAÇÃO 4.2. Não custa ressaltar que

$$V(\mathbb{T}, \lambda) = \ker(\mathbb{T} - \lambda I) \subset \bigcup_{k \in \mathbb{N}} \ker((\mathbb{T} - \lambda I)^k) = G(\mathbb{T}, \lambda).$$

O autoespaço generalizado é, também, um subespaço invariante pela transformação.

TEOREMA 4.1. *Seja  $\mathbb{T} \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  e suponha que  $\lambda$  seja um autovalor de  $\mathbb{T}$ . Então o autoespaço generalizado  $G(\mathbb{T}, \lambda) = \bigcup_{k \in \mathbb{N}} \ker(\mathbb{T} - \lambda I)^k$  é um subespaço invariante por  $\mathbb{T}$ . Além disso, se  $\mathbf{u} \in G(\mathbb{T}, \lambda)$  é um autovetor generalizado com índice  $k$ , então  $\mathbb{T}\mathbf{u}$  é um autovetor generalizado com índice no máximo  $k$ .*

DEMONSTRAÇÃO. Seja  $\mathbf{u} \in G(\mathbb{T}, \lambda)$ . Então existe  $k \in \mathbb{N}$  tal que  $(\mathbb{T} - \lambda I)^k \mathbf{u} = \mathbf{0}$ . Seja  $\mathbf{v} = \mathbb{T}\mathbf{u}$ . Então

$$(\mathbb{T} - \lambda I)^k \mathbf{v} = (\mathbb{T} - \lambda I)^k \mathbb{T}\mathbf{u} = \mathbb{T}(\mathbb{T} - \lambda I)^k \mathbf{u} = \mathbb{T}\mathbf{0} = \mathbf{0},$$

de modo que  $\mathbf{v} = \mathbb{T}\mathbf{u} \in G(\mathbb{T}, \lambda)$ , mostrando que  $G(\mathbb{T}, \lambda)$  é invariante por  $\mathbb{T}$ . Além disso,  $\mathbf{v}$  tem um índice no máximo igual ao de  $\mathbf{u}$ , como autovetor generalizado.  $\square$

OBSERVAÇÃO 4.3. No Teorema 4.1, o vetor  $\mathbf{v} = \mathbb{T}\mathbf{u}$  tem o mesmo índice  $k$  de  $\mathbf{u}$ , como autovalor generalizado, quando o autovalor  $\lambda \neq 0$ , e tem índice  $k - 1$ , quando  $\lambda = 0$ . Deixamos a verificação disso para o leitor, como Exercício 6.7.

EXEMPLO 4.2. Tomemos como exemplo o operador linear em  $\mathbb{R}^3$  associado à matriz

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 3 \\ -4 & 4 & 4 \\ 0 & 0 & 3 \end{bmatrix}$$

Como veremos no Exemplo 4.4, os autovalores são 2 e 3. Para  $\lambda = 3$ , o autoespaço é dado por

$$\begin{cases} y + 3z = 3x \\ -4x + 4y + 4z = 3y \\ 3z = 3z \end{cases}$$

cujas soluções são dadas por  $y = 0$  e  $z = x$ . Portanto, temos o autoespaço associado

$$V(\mathbf{A}, 3) = \{y = 0, z = x\} = \{(s, 0, s); s \in \mathbb{R}\}.$$

Para  $\lambda = 2$ , o autoespaço deve satisfazer é dado por

$$\begin{cases} y + 3z = 2x, \\ -4x + 4y + 4z = 2y, \\ 3z = 2z \end{cases}$$

Nesse caso,  $z = 0$  e o sistema se reduz a

$$\begin{cases} y = 2x, \\ -4x + 4y = 2y, \end{cases}$$

cuja solução é  $y = 2x$ . Portanto,

$$V(\mathbf{A}, 2) = \{(x, y, z); z = 0, y = 2x\} = \{(r, 2r, 0), r \in \mathbb{R}\}$$

Nesse caso, achamos dois autovalores, cada um com autoespaço de dimensão um. Ou seja, não há como termos uma base de autovalores para diagonalizar o operador. Vamos buscar um autoespaço generalizado de dimensão maior, para completar o espaço.

Vamos olhar para o núcleo de  $(\mathbf{A} - 3\mathbf{I})^2$ , que é dado por

$$(\mathbf{A} - 3\mathbf{I})^2 = \begin{bmatrix} 5 & -2 & -5 \\ 8 & -3 & -8 \\ 0 & 0 & 0 \end{bmatrix}$$

O núcleo é dado pelas soluções do sistema

$$\begin{cases} 5x - 2y - 5z = 0, \\ 8x - 3y - 8z = 0. \end{cases}$$

Multiplicando a primeira equação por 8, a segunda por 5 e subtraindo uma da outra, obtemos  $y = 0$  e  $z = x$ , que é o próprio autoespaço, ou seja

$$G(\mathbf{A}, 3) = V(\mathbf{A}, 3).$$

Vamos, agora, ver o autoespaço generalizado associado ao autovalor 2. Como o espaço tem dimensão três, esperamos que  $G(\mathbf{A}, 2)$  tenha dimensão dois, pois é o máximo que o espaço comporta dado que  $G(\mathbf{A}, 3) = V(\mathbf{A}, 3)$  tem dimensão um. Temos

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

O núcleo é dado pela solução do sistema associado, que nesse caso se reduz a  $z = 0$ . Logo,

$$G(\mathbf{A}, 2) = \ker((\mathbf{A} - 2\mathbf{I})^2) = \{z = 0\} = \{(x, y, 0); x, y \in \mathbb{R}\}.$$

OBSERVAÇÃO 4.4. Em dimensão finita, as potências  $(T - \lambda)^k$  devem se anular para algum  $k$  finito, de modo que o operador  $(T - \lambda)$  restrito ao autoespaço generalizado  $G(T, \lambda)$  é um operador nilpotente e existe um inteiro positivo  $k \in \mathbb{N}$  tal que

$$\{\mathbf{0}\} \subsetneq \ker(T - \lambda) \subsetneq \cdots \subsetneq \ker(T - \lambda)^k = G(T, \lambda).$$

OBSERVAÇÃO 4.5. Em dimensão infinita, podemos ter um autoespaço generalizado também de dimensão infinita e com espaços estritamente crescentes  $\ker(T - \lambda)^k$ ,  $k \in \mathbb{N}$ . De fato, considere  $X = \mathcal{C}(\mathbb{R})$  o espaço das funções infinitas vezes continuamente diferenciáveis e seja  $D = d/dt$  o operador derivada de uma função escalar  $x = x(t)$ . O autoespaço de  $D$  associado a um  $\lambda \in \mathbb{R}$  qualquer é a solução geral da equação diferencial

$$\frac{dx}{dt} = \lambda x,$$

ou seja,

$$V(D, \lambda) = \{x \in X; x(t) = C_0 e^{\lambda t}\}.$$

Agora, para um  $k \in \mathbb{N}$  qualquer, temos que  $x(t) = t^{k-1} e^{\lambda t}$  pertence ao núcleo de  $(D - \lambda)^k$ . Mais precisamente, podemos escrever

$$\ker((D - \lambda)^k) = \{(a_0 + a_1 t + \dots + a_{k-1} t^{k-1}) e^{\lambda t}; a_0, \dots, a_{k-1} \in \mathbb{R}\},$$

para  $k \in \mathbb{N}$  arbitrário. Com isso,

$$\begin{aligned} G(T, \lambda) &= \bigcup_{k \in \mathbb{N}} \ker((D - \lambda)^k) \\ &= \{(a_0 + a_1 t + \dots + a_{k-1} t^{k-1}) e^{\lambda t}; k \in \mathbb{N}, a_0, \dots, a_{k-1} \in \mathbb{R}\}. \end{aligned}$$

Veja, também, o [Exemplo 1.5](#).

O resultado do [Teorema 2.1](#) se estende para autoespaços generalizados.

TEOREMA 4.2. *Seja  $T \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  sobre um corpo  $\mathbb{K}$ . Sejam  $\lambda, \mu \in \mathbb{K}$  autovalores distintos de  $T$ . Então*

$$G(T, \lambda) \cap G(T, \mu) = \{\mathbf{0}\},$$

onde  $G(T, \lambda) = \bigcup_{k \in \mathbb{N}} \ker(T - \lambda)^k$  e  $G(T, \mu) = \bigcup_{k \in \mathbb{N}} \ker(T - \mu)^k$  são os autoespaços generalizados associados a  $\lambda$  e  $\mu$ .

DEMONSTRAÇÃO. Suponha que  $\mathbf{u} \in G(T, \lambda) \cap G(T, \mu)$ . Sejam  $k, l$  os menores inteiros não negativos tais que  $(T - \lambda)^k \mathbf{u} = \mathbf{0}$  e  $(T - \mu)^l \mathbf{u} = \mathbf{0}$ .

Suponhamos, primeiro, que  $k, l \geq 1$ . Sem perda de generalidade, suponha que  $l \geq k$ . Defina

$$\mathbf{v} = (T - \lambda)^{k-1} \mathbf{u} \neq \mathbf{0},$$

que é não nulo pois  $k$  é a menor potência que anula  $\mathbf{u}$ . Nesse caso, temos que  $\mathbf{v}$  é um autovetor associado ao autovalor  $\lambda$ , pois

$$(\mathbb{T} - \lambda)\mathbf{v} = (\mathbb{T} - \lambda)(\mathbb{T} - \lambda)^{k-1}\mathbf{u} = (\mathbb{T} - \lambda)^k\mathbf{u} = \mathbf{0}.$$

Agora, como  $(\mathbb{T} - \lambda)$  e  $(\mathbb{T} - \mu)$  comutam, também temos

$$(\mathbb{T} - \mu)^l\mathbf{v} = (\mathbb{T} - \mu)^l(\mathbb{T} - \lambda)^{k-1}\mathbf{u} = (\mathbb{T} - \lambda)^{k-1}(\mathbb{T} - \mu)^l\mathbf{u} = \mathbf{0}.$$

Por outro lado, também podemos escrever, usando a fórmula do binômio de Newton,

$$\mathbf{0} = (\mathbb{T} - \mu)^l\mathbf{v} = (\mathbb{T} - \lambda + (\lambda - \mu)\mathbb{I})^l\mathbf{v} = \sum_{j=0}^l \binom{l}{j} (\mathbb{T} - \lambda)^j (\lambda - \mu)^{l-j}\mathbf{v}.$$

Como  $(\mathbb{T} - \lambda)\mathbf{v} = \mathbf{0}$ , todas os termos se anulam, exceto  $j = 0$ , o que nos dá

$$\mathbf{0} = (\mathbb{T} - \mu)^l\mathbf{v} = (\mu - \lambda)^l\mathbf{v}.$$

Como  $\mu \neq \lambda$ , isso implica em  $\mathbf{v} = \mathbf{0}$ , o que significa

$$\mathbf{v} = (\mathbb{T} - \lambda)^{k-1}\mathbf{u} = \mathbf{0},$$

contradizendo a escolha de  $k$ .

Devemos ter, então,  $k = 0$  ou  $l = 0$ . Suponha que  $k = 0$ . Então  $(\mathbb{T} - \lambda)^k = (\mathbb{T} - \lambda)^0 = \mathbb{I}$ , o que significa que  $\mathbf{u} = \mathbf{0}$ . Idem se  $l = 0$ .

Em resumo, o caso  $k, l \geq 1$  nos dá uma contradição, de modo que devemos ter  $k = 0$  e/ou  $l = 0$ , caso este em que  $\mathbf{u} = \mathbf{0}$ , provando que  $G(\mathbb{T}, \lambda) \cap G(\mathbb{T}, \mu) = \{\mathbf{0}\}$ .  $\square$

Mais importante é o caso de vários autoespaços generalizados.

**TEOREMA 4.3.** *Seja  $\mathbb{T} \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  sobre um corpo  $\mathbb{K}$ . Sejam  $\lambda_1, \dots, \lambda_k \in \mathbb{K}$  autovalores distintos de  $\mathbb{T}$ . Então os autoespaços generalizados  $G(\mathbb{T}, \lambda_j) = \cup_{l \in \mathbb{N}} \ker(\mathbb{T} - \lambda_j \mathbb{I})^l$  são autoespaços independentes entre si, de modo que vale a soma direta*

$$G(\mathbb{T}, \lambda_1) \oplus \cdots \oplus G(\mathbb{T}, \lambda_k).$$

**DEMONSTRAÇÃO.** Graças ao **Teorema 2.1**, basta mostrar que, se  $\mathbf{u}_j \in G(\mathbb{T}, \lambda_j)$ ,  $j = 1, \dots, k$ , com

$$\mathbf{u}_1 + \mathbf{u}_2 + \cdots + \mathbf{u}_j = \mathbf{0}, \tag{4.1}$$

então cada  $\mathbf{u}_j = \mathbf{0}$  é nulo. Como de hábito, podemos mostrar isso por indução em  $k$ . Se  $k = 1$ , só temos um autoespaço generalizado e a condição (4.1) se reduz a  $\mathbf{u}_1 = \mathbf{0}$ , que nos diz diretamente que  $\mathbf{u}_1$  deve ser nulo.

Agora, supondo o resultado válido até  $k - 1$ , basta aplicar  $(T - \lambda_1 I)^{k_1}$  à expressão (4.1), onde  $k_1$  é o índice do autovetor generalizado  $\mathbf{u}_1$ , de modo que

$$\mathbf{0} + (T - \lambda_1 I)^{k_1} \mathbf{u}_2 + \cdots + (T - \lambda_1 I)^{k_1} \mathbf{u}_k = \mathbf{0}.$$

Como os autoespaços generalizados são invariantes, cada  $(T - \lambda_1 I)^{k_1} \mathbf{u}_j$  está em  $G(T, \lambda_j)$ . Portanto, temos  $k - 1$  vetores, cada um em um autoespaço generalizado diferente, e cuja soma é igual ao vetor nulo. Pela hipótese de indução, isso significa que

$$(T - \lambda_1 I)^{k_1} \mathbf{u}_j = \mathbf{0},$$

para  $j = 2, \dots, k$ , ou seja  $\mathbf{u}_j \in G(T, \lambda_1)$  também pertence ao autoespaço generalizado associado a  $\lambda_1$ , além de já pertencer a  $G(T, \lambda_j)$ .

Como os autoespaços generalizados tem interseção trivial, conforme mostrado no Teorema 4.2, então deduzimos que  $\mathbf{u}_j = \mathbf{0}$ , para  $j = 2, \dots, k$ . Com isso, (4.1) nos dá, como condição restante,  $\mathbf{u}_1 = \mathbf{0}$ , completando a demonstração.  $\square$

**4.2. Polinômios anuladores.** Na demonstração de existência de autovalores de um operador linear  $T$  em um espaço  $X$  de dimensão finita  $n \in \mathbb{N}$ , dado um vetor não nulo  $\mathbf{u}$ , construímos um polinômio  $p = p_{\mathbf{u}}$  de grau  $n$  tal que  $p_{\mathbf{u}}(T)\mathbf{u} = \mathbf{0}$ . A importância desse polinômio foi a de nos dar um autovalor como uma de suas raízes. Mas esse polinômio pode variar com a escolha de  $\mathbf{u}$  e podemos não encontrar todos os autovalores de uma só vez. Vamos rever um dos exemplos com mais diligência.

EXEMPLO 4.3. No exemplo  $T(x, y) = (2x, 3y)$ , se começarmos com  $\mathbf{u}$  no eixo  $x$ , i.e.  $\mathbf{u} = (a, 0)$  para  $a \neq 0$ , então pode-se verificar que  $p_{(a,0)}(x) = (x - 2)$  e que  $p_{(a,0)}(T)$  só anula o autoespaço  $V(T, 2)$ . Se começarmos no eixo  $y$ , digamos  $\mathbf{u} = (0, b)$ , com  $b \neq 0$ , então  $p_{(0,b)}(x) = (x - 3)$  e  $p_{(0,b)}(T)$  só anula o autoespaço  $V(T, 3)$ . Já se começarmos com  $\mathbf{u} = (a, b)$ ,  $a, b \neq 0$ , então  $p_{(a,b)} = (x - 2)(x - 3)$  e  $p_{(a,b)}(T)$  anula, finalmente, o espaço todo.

A nossa busca é por um polinômio que anule todo o espaço e nos dê, assim, todos os autovalores do operador, além de ser importante para contabilizar as dimensões do espaço entre as dimensões dos autoespaços ou autoespaços generalizados do operador. Nesse sentido, fazemos a seguinte definição.

DEFINIÇÃO 4.2. Dado um operador linear  $T \in \mathcal{L}(X)$  em um espaço vetorial  $X$  sobre um corpo  $\mathbb{K}$ , um polinômio  $p$  sobre o corpo  $\mathbb{K}$  é dito um **polinômio anulador** quando  $p(T) = 0$ , i.e.  $p(T)\mathbf{u} = \mathbf{0}$ , para todo  $\mathbf{u} \in X$ .

Um tal polinômio será fundamental para a decomposição espectral e o entendimento de operadores lineares. Uma questão crucial é como encontrar esses polinômios anuladores. Veremos, posteriormente, que o polinômio característico, obtido através

do determinante, é um polinômio anulador. No momento, vamos usar a forma triangular superior para obter um polinômio anulador. Posteriormente, vamos ver que esse polinômio obtido através da forma triangular é o próprio polinômio característico.

**TEOREMA 4.4.** *Seja  $T \in \mathcal{L}(X)$  um operador linear em um espaço vetorial complexo  $X$  de dimensão finita  $n \in \mathbb{N}$ . Suponha que  $\mathcal{b}$  seja uma base na qual a representação de  $T$  nessa base tenha forma triangular superior. Então o polinômio*

$$p(z) = (z - a_{11}) \cdots (z - a_{nn})$$

*cujas raízes  $a_{jj}$ ,  $j = 1, \dots, n$  formam a diagonal de  $[T]_{\mathcal{b}}$ , é um polinômio anulador de  $T$ , formado apenas por raízes que são autovalores do operador.*

**DEMONSTRAÇÃO.** Seja  $\mathcal{b} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  uma base como considerada no enunciado e seja  $[T]_{\mathcal{b}} = (a_{ij})_{ij}$  a representação de  $T$  na base  $\mathcal{b}$ , de forma que  $a_{ij} = 0$ , para  $i > j$ . Já vimos, no **Teorema 3.4**, que esses elementos são autovalores, de modo que todas as raízes de  $p$  são autovalores de  $T$ . Resta mostrar que  $p(T)$  anula  $X$ .

Para simplificar a notação, escrevemos

$$S_j = \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_j\},$$

com  $S_n = X$  e  $S_0 = \{\mathbf{0}\}$ . A base considerada tem a propriedade de que

$$TS_j \subset S_j, \quad j = 1, \dots, n.$$

Mais explicitamente,

$$T\mathbf{w}_j = a_{1j}\mathbf{w}_1 + \dots + a_{jj}\mathbf{w}_j,$$

de modo que

$$(T - a_{jj}I)\mathbf{w}_j = a_{1j}\mathbf{w}_1 + \dots + a_{(j-1)j}\mathbf{w}_{j-1} \in S_{j-1},$$

inclusive quando  $j = 1$ . Além disso,

$$(T - a_{jj}I)\mathbf{w}_i \in S_i \subset S_{j-1},$$

para todo  $i = 1, \dots, j - 1$ . Logo,

$$(T - a_{jj}I)S_j \subset S_{j-1},$$

para  $j = 1, \dots, n$ . Assim, temos

$$\begin{aligned} p(T)X &= (T - a_{11}I) \cdots (T - a_{nn}I)S_n \subset (T - a_{11}I) \cdots (T - a_{(n-1)(n-1)}I)S_{n-1} \\ &\subset \cdots \subset (T - a_{11}I)S_1 \subset S_0 = \{\mathbf{0}\}, \end{aligned}$$

ou seja,  $p(T) = 0$ , concluindo a demonstração.  $\square$

OBSERVAÇÃO 4.6. Uma outra maneira de ver que  $p(\mathbb{T}) = 0$  é via a representação matricial dos operadores envolvidos. Desse modo, vemos que cada monômio

$$\mathbb{T} - a_{jj}\mathbb{I}$$

é representado na base  $\mathcal{B}$  com um elemento nulo da linha  $j$  e coluna  $j$ . Assim,

$$[p(\mathbb{T})]_{\mathcal{B}} = [\mathbb{T} - a_{11}\mathbb{I}]_{\mathcal{B}} \cdots [\mathbb{T} - a_{nn}\mathbb{I}]_{\mathcal{B}}$$

é um produto de matrizes com um elemento nulo na linha  $j$  e coluna  $j$  correspondente, com  $j = 1, \dots, n$ . A cada multiplicação da direita para a esquerda, obtemos uma matriz com mais uma linha inferior nula. A cada multiplicação, anulamos uma nova linha, até chegarmos em toda a matriz  $[p(\mathbb{T})]_{\mathcal{B}}$  sendo nula. Também podemos fazer isso da esquerda para a direita, anulando sucessivamente as colunas, também da esquerda para a direita.

OBSERVAÇÃO 4.7. Um polinômio anulador não é único. De fato, qualquer monômio  $p(z) = (z - 1)^k$ ,  $k \in \mathbb{N}$  anula o operador identidade, visto que  $p(\mathbb{I}) = (\mathbb{I} - \mathbb{I})^k = 0$ . Podemos, também, multiplicar por qualquer outro monômio que o produto continua sendo anulador, e.g.  $p(z) = (z - 1)z$ . Observe que nem mesmo se nos restringirmos a polinômios do mesmo grau do espaço isso define unicamente o polinômio anulador. E nem mesmo se nos restringirmos a polinômios com raízes sendo autovalores. De fato, considerando o operador  $\mathbb{T}(x, y, z) = (2x, 2y, 3z)$ , os polinômios

$$p_1(z) = (z - 2)^2(z - 3), \quad p_2(z) = (z - 2)(z - 3)^2$$

ambos anulam  $\mathbb{T}$ . A chave está na dimensão dos núcleos das potências  $(\mathbb{T} - \lambda_j\mathbb{I})^k$ , como veremos a seguir.

O polinômio construído no Teorema 4.4 é exatamente o *polinômio característico* visto nos cursos iniciais de Álgebra Linear.

DEFINIÇÃO 4.3. Seja  $\mathbb{T} \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  de dimensão finita. O **polinômio característico** de  $\mathbb{T}$  é o polinômio dado por

$$p_c(z) = \det(z\mathbb{I} - \mathbb{T}).$$

OBSERVAÇÃO 4.8. Graças à fórmula de Leibniz (3.1), podemos verificar que  $p_c(z)$  é um polinômio mônico de grau exatamente  $n$ , lembrando que um polinômio mônico é um polinômio cujo coeficiente do monômio de maior grau é 1, ou seja, é um polinômio da forma  $z^n + q(z)$ , onde  $n \in \mathbb{N}$  e  $q = q(z)$  é um polinômio qualquer de grau estritamente menor do que  $n$ .

**TEOREMA 4.5.** *Seja  $T \in \mathcal{L}(X)$  um operador linear em um espaço vetorial complexo  $X$  de dimensão finita  $n \in \mathbb{N}$ . Suponha que  $\mathcal{B}$  seja uma base na qual a representação de  $T$  nessa base tenha forma triangular superior. Então o polinômio característico de  $T$  é exatamente o polinômio definido no [Teorema 4.4](#).*

**DEMONSTRAÇÃO.** A definição do determinante de um operador se baseia no fato de que o determinante de matrizes é invariante por similaridade. Dessa forma, definimos o determinante de um operador através do determinante da matriz que representa o operador em uma base arbitrária. Nesse sentido, temos que

$$p_c(z) = \det(zI - T) = \det([zI - T]_{\mathcal{B}}) = \det(zI - [T]_{\mathcal{B}}).$$

Como  $zI$  é diagonal e  $[T]_{\mathcal{B}}$  é triangular superior, então  $zI - [T]_{\mathcal{B}}$  é triangular superior e cuja diagonal tem a forma  $z - a_{jj}$ ,  $j = 1, \dots, n$ . De acordo com [Observação 3.5](#), o determinante de uma matriz triangular superior é o produto dos elementos da diagonal. Logo,

$$p_c(z) = (z - a_{11}) \cdots (z - a_{nn}),$$

coincidindo com o polinômio construído no [Teorema 4.4](#).  $\square$

Como o polinômio em [Teorema 4.4](#) é um polinômio anulador e coincide com o polinômio característico, então o polinômio característico é um polinômio anulador. Esse é um importante resultado que leva o nome de *Teorema de Cayley-Hamilton*. Há várias demonstrações desse teorema, algumas mais algébricas do que outras. A demonstração apresentada aqui explora a estrutura de subespaços invariantes encaixantes.

**COROLÁRIO 4.1** (Teorema de Cayley-Hamilton). *Seja  $T \in \mathcal{L}(X)$  um operador linear em um espaço vetorial complexo  $X$  de dimensão finita  $n \in \mathbb{N}$ . Então o polinômio característico  $p_c = p_c(z)$  de  $T$  satisfaz  $p_c(T) = 0$  e, portanto, é um polinômio anulador de  $T$ .*

**OBSERVAÇÃO 4.9.** Observe que no caso de um operador  $T = \lambda I$  em  $\mathbb{R}^n$ , o seu polinômio característico tem a forma

$$p_c(z) = (z - \lambda)^n.$$

No entanto, basta uma potência para anular o operador, i.e.  $T - \lambda I = 0$ . Já no caso de um operador  $T = \lambda I + N$ , onde  $N$  é um operador nilpotente, então temos também

$$p_c(z) = (z - \lambda)^n,$$

então precisamos de uma potência igual ao índice de  $N$  para anular o operador. Isso nos leva ao conceito de *polinômio minimal*, que é o polinômio mônico de menor grau

possível que ainda anula o operador. Sendo assim, o polinômio minimal de  $T = \lambda I$  é

$$p_m(z) = z - \lambda,$$

enquanto que o polinômio minimal de  $T = \lambda I + N$ , onde  $N$  é um operador nilpotente de índice  $k$ , é dado por

$$p_c(z) = (z - \lambda)^k.$$

Mais geralmente, dado um operador qualquer  $T$  nos complexos, o seu polinômio característico tem a forma

$$p_c(z) = (z - \lambda_1)^{n_1} \cdots (z - \lambda_m)^{n_m},$$

onde  $\lambda_j = 1, \dots, m$ ,  $m \in \mathbb{N}$ ,  $1 \leq m \leq n$ , são autovalores distintos de  $T$  e  $n_j \in \mathbb{N}$  satisfazem

$$n_1 + \cdots + n_m = n.$$

Por sua vez, o polinômio minimal tem a forma

$$p_m(z) = (z - \lambda_1)^{k_1} \cdots (z - \lambda_m)^{k_m},$$

onde  $k_j \in \mathbb{N}$ ,  $1 \leq k_j \leq n_j$ , é exatamente o índice de nilpotência de  $(T - \lambda_j I)|_{G(T, \lambda_j)}$ , com  $G(T, \lambda_j) = \ker((T - \lambda_j I)^{k_j})$ .

**4.3. Multiplicidade dos autovalores e dimensão do autoespaço generalizado.** O próximo resultado é o resultado principal para obtermos a forma canônica de Jordan.

**TEOREMA 4.6.** *Seja  $T \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  de dimensão finita  $n \in \mathbb{N}$ . Suponha que  $\mathfrak{b}$  seja uma base na qual a representação de  $T$  nessa base tenha forma triangular superior e seja  $p$  o polinômio anulador dado por*

$$p(z) = (z - a_{11}) \cdots (z - a_{nn})$$

*cujas raízes formam a diagonal de  $[T]_{\mathfrak{b}} = (a_{ij})_{i,j=1}^{i,j=n}$  e são os autovalores de  $T$ . Juntando as diagonais de mesmo valor em um mesmo monômio, com o grau do monômio contando quantas vezes um mesmo autovalor aparece na diagonal, podemos escrever*

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

*onde  $\lambda_1, \dots, \lambda_k$  são autovalores distintos e  $n_1, \dots, n_k \in \mathbb{N}$  são as multiplicidades de cada autovalor. Então, cada autoespaço generalizado  $G(T, \lambda_j) = \cup_{m \in \mathbb{N}} (T - \lambda_j I)^m$ ,  $j = 1, \dots, k$ , é tal que*

$$\dim(G(T, \lambda_j)) = n_j \tag{4.2}$$

e

$$G(T, \lambda_j) = \ker(T - \lambda_j I)^{m_j}, \tag{4.3}$$

para um menor inteiro  $m_j$  possível, satisfazendo  $1 \leq m_j \leq n_j$ , com

$$n_1 + n_2 + \cdots + n_k = n. \quad (4.4)$$

Além disso, o espaço se decompõe em soma direta dos autoespaços generalizados:

$$X = G(\mathbb{T}, \lambda_1) \oplus G(\mathbb{T}, \lambda_2) \oplus \cdots \oplus G(\mathbb{T}, \lambda_k). \quad (4.5)$$

DEMONSTRAÇÃO. Como o polinômio  $p$  é de grau  $n$  e  $n_j$  é o grau de cada um de seus monômios, então é claro que (4.4) vale. Vamos agora mostrar (4.2). Vamos mostrar isso para um dado  $j$  arbitrário. Para isso, denotamos

$$S_i = \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_i\}, \quad (4.6)$$

para todo  $i = 1, \dots, n$ , com

$$S_0 = \{\mathbf{0}\}.$$

Pela natureza da base, que dá a  $[\mathbb{T}]_\theta$  uma forma triangular superior, vale

$$\mathbb{T}S_i \subset S_i,$$

tanto para  $i = 1, \dots, n$ , quanto para  $i = 0$ , com os dois conjuntos sendo triviais, i.e.  $\mathbb{T}S_0 = S_0 = \{\mathbf{0}\}$ . Com isso, também vale

$$(\mathbb{T} - \lambda\mathbb{I})S_i \subset S_i, \quad (4.7)$$

para qualquer  $\lambda$ , autovalor ou não.

O índice  $n_j$  indica quantas vezes o autovalor  $\lambda_j$  aparece na diagonal de  $[\mathbb{T}]_\theta$ . Com isso, existem índices  $i_1, \dots, i_{n_j}$ , ordenados de forma crescente, tais que  $a_{ii} = \lambda_j$ , para  $i = i_1, \dots, i_{n_j}$ . Vamos mostrar por indução que

$$\dim(\ker(\mathbb{T} - \lambda_j\mathbb{I})^l|_{S_{i_l}}) \geq l, \quad (4.8)$$

para  $l = 1, \dots, n_j$ .

Primeiramente, observe que

$$\mathbb{T}\mathbf{w}_i = a_{1i}\mathbf{w}_1 + \cdots + a_{ii}\mathbf{w}_i,$$

de modo que

$$(\mathbb{T} - \lambda_j\mathbb{I})\mathbf{w}_i = (\mathbb{T} - a_{ii}\mathbb{I})\mathbf{w}_i = a_{1i}\mathbf{w}_1 + \cdots + a_{(i-1)i}\mathbf{w}_{i-1} \in S_{i-1},$$

para todo  $i_1, \dots, i_{n_j}$ . Além disso, de (4.7), temos

$$(\mathbb{T} - \lambda_j\mathbb{I})S_{i-1} \subset S_{i-1},$$

para todo  $i = 1, \dots, n$ , em particular para  $i_1, \dots, i_{n_j}$ . Juntando as duas inclusões, temos

$$(\mathbb{T} - \lambda_j\mathbb{I})S_i \subset S_{i-1}, \quad (4.9)$$

para todo  $i_1, \dots, i_{n_j}$ . Portanto, olhando para  $(\mathbb{T} - \lambda_j \mathbb{I})$  restrito a  $S_i$ , que é denotado por  $(\mathbb{T} - \lambda_j \mathbb{I})|_{S_i}$ , temos, pelo **Teorema 2.1** do Núcleo e da Imagem,

$$\begin{aligned} \dim(\ker((\mathbb{T} - \lambda_j \mathbb{I})|_{S_i})) &= \dim(S_i) - \dim(\text{Im}((\mathbb{T} - \lambda_j \mathbb{I})|_{S_i})) \\ &\geq \dim(S_i) - \dim(S_{i-1}) = i - (i - 1) = 1, \end{aligned} \quad (4.10)$$

para todo  $i = i_1, \dots, i_{n_j}$ .

Em particular, para  $i = i_1$ , a desigualdade (4.10) nos dá

$$\dim(\ker(\mathbb{T} - \lambda_j \mathbb{I})|_{S_{i_1}}) \geq 1,$$

que é exatamente (4.8) para  $l = 1$ . Ou seja, já temos o primeiro passo no processo de indução para mostrar (4.8) para todo  $l = 1, \dots, n_j$ . Observe que esse primeiro passo é equivalente a mostrar que  $\lambda_j$  é um autovalor que tem pelo menos parte do seu autoespaço dentro de  $S_{i_1}$ , de dimensão pelo menos um.

Suponha, agora, (4.8) válido até  $l$ , onde  $1 \leq l < n_j$ . Vamos mostrar para  $l + 1$ . Consideramos  $\mathbb{T}$  restrito a  $S_{i_{l+1}}$ . Aplicando o **Teorema 2.1** do Núcleo e da Imagem, temos

$$\dim(\ker(\mathbb{T} - \lambda_j \mathbb{I})^{l+1}|_{S_{i_{l+1}}}) = \dim(S_{i_{l+1}}) - \dim(\text{Im}(\mathbb{T} - \lambda_j \mathbb{I})^{l+1}|_{S_{i_{l+1}}}).$$

Reescrevendo a imagem de  $(\mathbb{T} - \lambda_j \mathbb{I})^{l+1}|_{S_{i_{l+1}}}$  como  $(\mathbb{T} - \lambda_j \mathbb{I})^{l+1}S_{i_{l+1}}$ , temos

$$\begin{aligned} \dim(\ker(\mathbb{T} - \lambda_j \mathbb{I})^{l+1}|_{S_{i_{l+1}}}) &= \dim(S_{i_{l+1}}) - \dim(\text{Im}(\mathbb{T} - \lambda_j \mathbb{I})^{l+1}|_{S_{i_{l+1}}}) \\ &= \dim(S_{i_{l+1}}) - \dim((\mathbb{T} - \lambda_j \mathbb{I})^{l+1}S_{i_{l+1}}). \end{aligned}$$

Usando (4.9) para  $i = i_{l+1}$ , obtemos

$$\begin{aligned} \dim(\ker(\mathbb{T} - \lambda_j \mathbb{I})^{l+1}|_{S_{i_{l+1}}}) &= \dim(S_{i_{l+1}}) - \dim(\text{Im}(\mathbb{T} - \lambda_j \mathbb{I})^{l+1}|_{S_{i_{l+1}}}) \\ &= \dim(S_{i_{l+1}}) - \dim((\mathbb{T} - \lambda_j \mathbb{I})^{l+1}S_{i_{l+1}}) \\ &\geq \dim(S_{i_{l+1}}) - \dim((\mathbb{T} - \lambda_j \mathbb{I})^l S_{i_{l+1}-1}). \end{aligned}$$

Usando novamente o **Teorema 2.1** do Núcleo e da Imagem, chegamos a

$$\begin{aligned} &\dim(\ker(\mathbb{T} - \lambda_j \mathbb{I})^{l+1}|_{S_{i_{l+1}}}) \\ &= \dim(S_{i_{l+1}}) - \dim(\text{Im}(\mathbb{T} - \lambda_j \mathbb{I})^{l+1}|_{S_{i_{l+1}}}) \\ &= \dim(S_{i_{l+1}}) - \dim((\mathbb{T} - \lambda_j \mathbb{I})^{l+1}S_{i_{l+1}}) \\ &\geq \dim(S_{i_{l+1}}) - \dim((\mathbb{T} - \lambda_j \mathbb{I})^l S_{i_{l+1}-1}) \\ &= \dim(S_{i_{l+1}}) - \dim(S_{i_{l+1}-1}) + \dim(\ker(\mathbb{T} - \lambda_j \mathbb{I})^l|_{S_{i_{l+1}-1}}) \end{aligned}$$

No último termo, diminuindo o espaço de restrição do operador  $(T - \lambda_j I)^l$  de  $S_{i_{l+1}-1}$  para  $S_i \subset S_{i_{l+1}-1}$ , o núcleo pode apenas diminuir, de modo que

$$\begin{aligned}
& \dim(\ker(T - \lambda_j I)^{l+1}|_{S_{i_{l+1}}}) \\
&= \dim(S_{i_{l+1}}) - \dim(\operatorname{Im}(T - \lambda_j I)^{l+1}|_{S_{i_{l+1}}}) \\
&= \dim(S_{i_{l+1}}) - \dim((T - \lambda_j I)^{l+1}S_{i_{l+1}}) \\
&\geq \dim(S_{i_{l+1}}) - \dim((T - \lambda_j I)^l S_{i_{l+1}-1}) \\
&= \dim(S_{i_{l+1}}) - \dim(S_{i_{l+1}-1}) + \dim(\ker(T - \lambda_j I)^l|_{S_{i_{l+1}-1}}) \\
&\geq \dim(S_{i_{l+1}}) - \dim(S_{i_{l+1}-1}) + \dim(\ker(T - \lambda_j I)^l|_{S_i}).
\end{aligned}$$

Finalmente, usando a hipótese de indução, junto com o fato de que cada  $S_j$  tem dimensão  $j$ , obtemos

$$\begin{aligned}
& \dim(\ker(T - \lambda_j I)^{l+1}|_{S_{i_{l+1}}}) \\
&= \dim(S_{i_{l+1}}) - \dim(\operatorname{Im}(T - \lambda_j I)^{l+1}|_{S_{i_{l+1}}}) \\
&= \dim(S_{i_{l+1}}) - \dim((T - \lambda_j I)^{l+1}S_{i_{l+1}}) \\
&\geq \dim(S_{i_{l+1}}) - \dim((T - \lambda_j I)^l S_{i_{l+1}-1}) \\
&= \dim(S_{i_{l+1}}) - \dim(S_{i_{l+1}-1}) + \dim(\ker(T - \lambda_j I)^l|_{S_{i_{l+1}-1}}) \\
&\geq \dim(S_{i_{l+1}}) - \dim(S_{i_{l+1}-1}) + \dim(\ker(T - \lambda_j I)^l|_{S_i}) \\
&\geq i_{l+1} - (i_{l+1} - 1) + l \\
&= l + 1,
\end{aligned}$$

provando a indução e obtendo (4.8). No caso em que  $l = n_j$ , obtemos

$$\dim(\ker(T - \lambda_j I)^{n_j}|_{S_{i_{n_j}}}) \geq n_j. \quad (4.11)$$

Como  $j = 1, \dots, k$  é arbitrário, isso vale para todos esses índices  $j$ .

Agora, observamos, do Teorema 4.3, que os autoespaços generalizados são linearmente independentes e formam uma soma direta, dentro de  $X$ . Juntando isso com (4.11), obtemos

$$\begin{aligned}
& n_1 + n_2 + \dots + n_k \\
&\leq \dim(G(T, \lambda_1)) + \dim(G(T, \lambda_2)) + \dots + \dim(G(T, \lambda_k)) \leq n.
\end{aligned}$$

Como  $n_j$  são os graus dos monômios que formam o polinômio  $p$  e que, por sua vez, tem grau  $n$ , então esse “sanduíche” fecha completamente e nos dá que

$$\begin{aligned} n &= n_1 + n_2 + \cdots + n_k \\ &= \dim(G(\mathbb{T}, \lambda_1)) + \dim(G(\mathbb{T}, \lambda_2)) + \cdots + \dim(G(\mathbb{T}, \lambda_k)) = \dim(X). \end{aligned}$$

Isso também nos dá que (4.11) é, na verdade, uma igualdade, provando (4.2), e que  $X$  é soma direta dos autoespaços generalizados, provando (4.5).

Por fim, a igualdade em (4.2) garante que cada autoespaço generalizado  $G(\mathbb{T}, \lambda_j)$  não pode ter índice maior do que  $n_j$ , pois ele é um operador nilpotente quando restrito a  $G(\mathbb{T}, \lambda_j)$  e nenhum operador nilpotente pode ter dimensão maior do que a do (sub)espaço em que ele age (Corolário 1.1). Com isso, completamos a demonstração.  $\square$

**OBSERVAÇÃO 4.10.** Uma consequência do Teorema 4.6 é o de que as diagonais de quaisquer duas representações do operador  $\mathbb{T}$  em forma triangular superior devem ser uma permutação uma da outra. Isso porque o teorema associa a multiplicidade do autovalor na diagonal à dimensão do autoespaço generalizado, que independe da base. Isso resolve a questão levantada na Observação 3.7.

Vamos escrever a observação anterior como um corolário.

**COROLÁRIO 4.2.** *Seja  $\mathbb{T} \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  de dimensão finita  $n \in \mathbb{N}$ . Suponha que  $\mathfrak{a}$  e  $\mathfrak{b}$  sejam duas bases tais que as representações  $[\mathbb{T}]_{\mathfrak{a}}$  e  $[\mathbb{T}]_{\mathfrak{b}}$  tenham forma triangular superior. Então a diagonal de  $[\mathbb{T}]_{\mathfrak{a}}$  é uma permutação da diagonal de  $[\mathbb{T}]_{\mathfrak{b}}$ , ou seja, os autovalores aparecem nas diferentes diagonais com a mesma multiplicidade.*

**4.4. Multiplicidade algébrica e multiplicidade geométrica dos autovalores.** Inspirado pelos resultados da seção anterior, fazemos a seguinte definição.

**DEFINIÇÃO 4.4.** *Seja  $\mathbb{T} \in \mathcal{L}(X)$  um operador linear em um espaço de dimensão finita. Seja  $\lambda$  um autovalor de  $\mathbb{T}$ . Definimos a **multiplicidade geométrica** de  $\lambda$  como sendo a dimensão do seu autoespaço,*

$$n_g^\lambda = n_g(\mathbb{T}, \lambda) = \dim(V(\mathbb{T}, \lambda)) = \dim(\ker(\mathbb{T} - \lambda I)),$$

*e definimos a **multiplicidade algébrica** de  $\lambda$  como sendo a dimensão do seu autoespaço generalizado,*

$$n_a^\lambda = n_a(\mathbb{T}, \lambda) = \dim(G(\mathbb{T}, \lambda)) = \dim(\cup_{k \in \mathbb{N}} \ker(\mathbb{T} - \lambda I)^k).$$

Como é de se esperar, a relação entre  $n_a(\mathbb{T}, \lambda)$  e  $n_g(\mathbb{T}, \lambda)$  é fundamental.

TEOREMA 4.7. *Seja  $T \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  de dimensão finita  $n = \dim(X) \in \mathbb{N}$  sobre um corpo  $\mathbb{K}$  e seja  $\lambda \in \mathbb{K}$  um autovalor de  $T$ . Então  $n_a(T, \lambda) \geq n_g(T, \lambda) \geq 1$  e*

$$T|_{G(T, \lambda)} = \lambda I + N,$$

onde a identidade é em  $G(T, \lambda)$  e  $N \in \mathcal{L}(G(T, \lambda))$  é um operador nilpotente em  $G(T, \lambda)$ , com índice de nilpotência no máximo  $n_a(T, \lambda)$ . No caso particular em que  $n_a(T, \lambda) = n_g(T, \lambda)$ , então  $G(T, \lambda) = V(T, \lambda)$  e

$$T|_{G(T, \lambda)} = T|_{V(T, \lambda)} = \lambda I.$$

DEMONSTRAÇÃO. A desigualdade  $n_a(T, \lambda) \geq n_g(T, \lambda)$  segue imediatamente do fato de que  $V(T, \lambda) \subset G(T, \lambda)$  (veja [Observação 4.2](#)), e, como  $\lambda$  é autovalor, existe pelo menos um  $\mathbf{u} \in X$  não nulo em  $V(T, \lambda)$ , de modo que  $n_a(T, \lambda) \geq n_g(T, \lambda) \geq 1$ .

Agora, restrito a  $G(T, \lambda)$ , que é um subespaço, definimos

$$N = (T - \lambda I)|_{G(T, \lambda)},$$

que é um operador linear em  $G(T, \lambda)$  (veja [Definição 5.2](#)).

Como  $G(T, \lambda)$  tem dimensão finita  $n_a^\lambda = n_a(T, \lambda)$ , então, escolhendo uma base  $\{\mathbf{w}_1, \dots, \mathbf{w}_{n_a^\lambda}\}$ , temos cada vetor da base com um determinado índice (como autovetor generalizado, dado na [Definição 4.1](#))  $k_i$ ,  $i = 1, \dots, n_a^\lambda$  de  $N = T - \lambda I$ , que é o menor inteiro tal que  $N^{k_i} \mathbf{w}_i = \mathbf{0}$ . Escolhendo  $k = \max\{k_1, \dots, k_{n_a^\lambda}\}$ , vemos que

$$N^k \mathbf{u} = \mathbf{0}, \quad \forall \mathbf{u} \in \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_{n_a^\lambda}\} = G(T, \lambda).$$

Portanto,  $N$  é um operador nilpotente em  $G(T, \lambda)$ . Como operador nilpotente, o seu índice deve ser no máximo a dimensão do espaço em que está definido, ou seja, o seu índice de nilpotência é no máximo  $n_a(T, \lambda)$ .

Agora, falta ver o caso em que  $n_a(T, \lambda) = n_g(T, \lambda)$ . Nesse caso, como  $V(T, \lambda) \subset G(T, \lambda)$  e ambos tem a mesma dimensão, então

$$G(T, \lambda) = V(T, \lambda).$$

Além disso, para todo  $\mathbf{u} \in G(T, \lambda) = V(T, \lambda)$ , temos  $T\mathbf{u} = \lambda\mathbf{u}$ , ou seja,

$$T|_{V(T, \lambda)} = \lambda I.$$

Isso completa a demonstração. □

**4.5. Forma canônica de Jordan.** Com os resultados das últimas seções, estamos prontos para juntar as peças e provar o principal resultado desse capítulo.

DEFINIÇÃO 4.5 (Forma canônica de Jordan). *Dizemos que uma matriz  $\mathbf{A} \in \mathbb{K}^{n \times n}$ ,  $n \in \mathbb{N}$ , está na **forma canônica de Jordan** quando  $\mathbf{A}$  é bloco-diagonal, i.e.*

$$\mathbf{A} = \begin{bmatrix} \mathbf{J}_1 & & & \\ & \mathbf{J}_2 & & \\ & & \ddots & \\ & & & \mathbf{J}_r \end{bmatrix},$$

onde cada bloco  $\mathbf{J}_l$ ,  $l = 1, \dots, r$ , é da forma  $\lambda_l \mathbf{I} + \mathbf{N}_l$ , onde  $\lambda_l \in \mathbb{K}$  é um autovalor de  $\mathbf{A}$  e  $\mathbf{N}_l = (\delta_{i(j-1)})_{ij}$  é um bloco com 1 na superdiagonal e zero no resto, de modo que

$$\mathbf{J}_l = \lambda_l \mathbf{I} + \mathbf{N}_l = \begin{bmatrix} \lambda_l & 1 & & \\ & \lambda_l & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_l \end{bmatrix}.$$

TEOREMA 4.8 (Forma canônica de Jordan nos complexos). *Seja  $\mathbf{T} \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  de dimensão finita  $n = \dim(X) \in \mathbb{N}$  sobre  $\mathbb{C}$ . Então existe uma base  $\mathfrak{b}$  de  $X$  em que  $[\mathbf{T}]_{\mathfrak{b}}$  está na forma canônica de Jordan.*

DEMONSTRAÇÃO. No corpo dos complexos, vimos, no Teorema 3.3, que existe uma base em que  $\mathbf{T}$  é triangular superior. A partir dessa base, deduzimos, do Teorema 4.6, que existem  $m$  autovalores distintos,  $\lambda_1, \dots, \lambda_m \in \mathbb{C}$ ,  $m \in \mathbb{N}$ , tais que

$$X = G(\mathbf{T}, \lambda_1) \oplus \dots \oplus G(\mathbf{T}, \lambda_m).$$

Pelo Teorema 4.7, temos

$$\mathbf{T}|_{G(\mathbf{T}, \lambda_j)} = \lambda_j \mathbf{I} + \mathbf{N}_j,$$

onde

$$\mathbf{N}_j = (\mathbf{T} - \lambda_j \mathbf{I})|_{G(\mathbf{T}, \lambda_j)} \in \mathcal{L}(G(\mathbf{T}, \lambda_j))$$

é nilpotente com índice  $1 \leq k_j \leq n_a(\mathbf{T}, \lambda_j)$ .

De acordo com o Teorema 3.2, existe uma base  $\mathfrak{b}_j$  de  $G(\mathbf{T}, \lambda_j)$  na qual  $\mathbf{N}_j$  está na forma canônica

$$[\mathbf{N}]_{\mathfrak{b}_j} = \begin{bmatrix} \mathbf{N}_{j,1} & & & \\ & \mathbf{N}_{j,2} & & \\ & & \ddots & \\ & & & \mathbf{N}_{j,r_j} \end{bmatrix}.$$

com blocos  $\mathbf{J}_{j,i}$  da forma

$$\mathbf{N}_{j,i} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}.$$

Nessa base,  $[\mathbf{T}|_{G(\mathbb{T}, \lambda_j)}]_{\mathfrak{b}_j}$  está na forma canônica de Jordan,

$$[\mathbf{T}|_{G(\mathbb{T}, \lambda_j)}]_{\mathfrak{b}_j} = \begin{bmatrix} \mathbf{J}_{j,1} & & & \\ & \mathbf{J}_{j,2} & & \\ & & \ddots & \\ & & & \mathbf{J}_{j,r_j} \end{bmatrix},$$

onde cada bloco  $\mathbf{J}_{j,i}$  é da forma

$$\mathbf{J}_{j,i} = \lambda_j \mathbf{I} + \mathbf{N}_{j,i} = \begin{bmatrix} \lambda_j & 1 & & \\ & \lambda_j & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_j \end{bmatrix}.$$

Juntando as bases de cada autoespaço generalizado  $G(\mathbb{T}, \lambda_j)$ , obtemos uma base

$$\mathfrak{b} = \mathfrak{b}_1 \cup \dots \cup \mathfrak{b}_m$$

de  $X$  na qual  $[\mathbf{T}]_{\mathfrak{b}}$  está na forma de Jordan

$$[\mathbf{T}]_{\mathfrak{b}} = \begin{bmatrix} \mathbf{J}_{1,1} & & & & & & & \\ & \ddots & & & & & & \\ & & \mathbf{J}_{1,r_1} & & & & & \\ & & & \mathbf{J}_{2,1} & & & & \\ & & & & \ddots & & & \\ & & & & & \mathbf{J}_{2,r_2} & & \\ & & & & & & \ddots & \\ & & & & & & & \mathbf{J}_{m,1} \\ & & & & & & & & \ddots & \\ & & & & & & & & & \mathbf{J}_{m,r_m} \end{bmatrix},$$

completando a demonstração. □

OBSERVAÇÃO 4.11. A forma canônica de Jordan, garantida pelo [Teorema 4.8](#), pode não ser única, visto que podemos reordenar a base  $e$ , com isso, reordenar os blocos de Jordan. De fato, o operador  $T(x, y, z, w) = (y, z, 0, 0)$  tem a seguinte forma canônica de Jordan na base canônica  $e = \{e_1, e_2, e_3, e_4\}$ ,

$$[T]_e = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

enquanto que, na base reordenada  $\ell = \{e_4, e_1, e_2, e_3\}$ ,

$$[T]_\ell = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

OBSERVAÇÃO 4.12. Continuando a [Observação 4.11](#), podemos fixar uma forma única ordenando os autovalores e ordenando os tamanhos dos blocos, digamos, ambos do menor para o maior, primeiro entre os autovalores e depois entre os tamanhos da cada bloco associados ao mesmo autovalor. Dessa maneira, a forma  $[T]_\ell$  na [Observação 4.11](#) seria a forma ordenada única. Mesmo assim, as bases podem não ser únicas. De fato, escolhendo outros ciclos podemos obter a mesma forma. Por exemplo, escolhendo

$$\alpha = \{(1, 0, 0, 0), (1, 1, 0, 0), (1, 1, 1, 0), (0, 0, 0, 1)\},$$

obtemos  $[T]_\alpha = [T]_\ell$ .

#### 4.6. Exemplos.

EXEMPLO 4.4. Tomemos como exemplo o operador linear em  $\mathbb{R}^3$  associado à matriz

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 3 \\ -4 & 4 & 4 \\ 0 & 0 & 3 \end{bmatrix}$$

Vamos, inicialmente, procurar o núcleo de  $\mathbf{A} - \lambda\mathbf{I}$ . Para isso, temos que resolver o sistema

$$\begin{cases} -\lambda x + y + 3z = 0, \\ -4x + (4 - \lambda)y + 4z = 0, \\ (3 - \lambda)z = 0. \end{cases}$$

Para  $\lambda = 3$ , a última linha se anula e devemos ter

$$\begin{cases} -3x + y + 3z = 0, \\ -4x + y + 4z = 0, \end{cases}$$

cujas soluções são dadas por  $y = 0$  e  $z = x$ . Portanto, obtemos um autovalor  $\lambda = 3$ , com autoespaço associado

$$V(\mathbf{A}, 3) = \{y = 0, z = x\} = \{(s, 0, s); s \in \mathbb{R}\}.$$

Para  $\lambda \neq 3$ , devemos ter  $z = 0$ , sobrando o sistema

$$\begin{cases} -\lambda x + y = 0, \\ -4x + (4 - \lambda)y = 0. \end{cases}$$

Fazendo  $y = \lambda x$ , obtemos a equação

$$-4x + (4 - \lambda)\lambda x = (\lambda^2 + 4\lambda - 4)x = 0.$$

A única solução não nula é quando

$$\lambda^2 + 4\lambda - 4 = (\lambda - 2)^2 = 0.$$

que nos dá uma raiz dupla,  $\lambda = 2$ . Com isso, temos  $\lambda = 2$  como o único outro autovalor. O autoespaço deve satisfazer  $z = 0$  e

$$\begin{cases} -2x + y = 0, \\ -4x + 2y = 0, \end{cases}$$

cuja solução é  $y = 2x$ . Portanto,

$$V(\mathbf{A}, 2) = \{(x, y, z); z = 0, y = 2x\} = \{(r, 2r, 0), r \in \mathbb{R}\}$$

Nesse caso, achamos dois autovalores, cada um com autoespaço de dimensão um. Ou seja, não há como termos uma base de autovalores para diagonalizar o operador. Vamos buscar um autoespaço generalizado de dimensão maior, para completar o espaço.

Vamos olhar para o núcleo de  $(\mathbf{A} - 3\mathbf{I})^2$ , que é dado por

$$(\mathbf{A} - 3\mathbf{I})^2 = \begin{bmatrix} 5 & -2 & -5 \\ 8 & -3 & -8 \\ 0 & 0 & 0 \end{bmatrix}$$

O núcleo é dado pelas soluções do sistema

$$\begin{cases} 5x - 2y - 5z = 0, \\ 8x - 3y - 8z = 0. \end{cases}$$

Multiplicando a primeira equação por 8, a segunda por 5 e subtraindo uma da outra, obtemos  $y = 0$  e  $z = x$ , que é o próprio autoespaço, ou seja

$$G(\mathbf{A}, 3) = V(\mathbf{A}, 3).$$

Vamos, agora, ver o autoespaço generalizado associado ao autovalor 2. Como o espaço tem dimensão três, esperamos que  $G(\mathbf{A}, 2)$  tenha dimensão dois, pois é o máximo que o espaço comporta dado que  $G(\mathbf{A}, 3) = V(\mathbf{A}, 3)$  tem dimensão um. Temos

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

O núcleo é dado pela solução do sistema associado, que nesse caso se reduz a  $z = 0$ . Logo,

$$G(\mathbf{A}, 2) = \ker((\mathbf{A} - 2\mathbf{I})^2) = \{z = 0\} = \{(x, y, 0); x, y \in \mathbb{R}\}.$$

Assim, vemos que

$$X = G(\mathbf{T}, 2) \oplus V(\mathbf{T}, 3).$$

Ao tomarmos uma base arbitrária de  $G(\mathbf{A}, 2)$  e uma base de  $V(\mathbf{A}, 3)$ , vamos conseguir uma forma diagonal em blocos, mas não necessariamente na forma canônica. Vejamos, só por experiência. Observe que, na base canônica, já temos uma forma triangular em blocos. Para a forma diagonal, podemos tomar a base

$$\mathfrak{a} = \{(1, 0, 0), (0, 1, 0), (1, 0, 1)\}$$

Nesse caso, a matriz mudança de base, junto com a sua inversa, é

$$\tilde{\mathbf{M}} = [I]_{\mathfrak{a}}^e = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \tilde{\mathbf{M}}^{-1} = [I]_e^{\mathfrak{a}} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

e temos

$$\tilde{\mathbf{M}}^{-1}\mathbf{A}\tilde{\mathbf{M}} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 3 \\ -4 & 4 & 4 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -4 & 4 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

Para a forma canônica, precisamos achar um  $\mathbf{u} \in \ker((\mathbf{A} - 2\mathbf{I})^2) \setminus \ker(\mathbf{A} - 2\mathbf{I})$ , por exemplo,

$$\mathbf{w}_2 = (1, 3, 0).$$

Este vetor nos gera um ciclo de dois elementos do operador nilpotente  $(\mathbf{T} - 2\mathbf{I})$ , com

$$\mathbf{w}_1 = (\mathbf{T} - 2\mathbf{I})(1, 3, 0) = (1, 2, 0) \in V(\mathbf{A}, 2).$$

Completamos com o mesmo autovetor do autovalor 3,

$$\mathbf{w}_3 = (1, 0, 1).$$

Isso nos dá a base

$$\mathcal{B} = \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\} = \{(1, 2, 0), (1, 3, 0), (1, 0, 1)\}.$$

A matriz mudança de base e a sua inversão são

$$\mathbf{M} = [I]_{\mathcal{B}}^e = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{M}^{-1} = [I]_e^{\mathcal{B}} = \begin{bmatrix} 3 & -1 & -3 \\ -2 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix},$$

e assim obtemos a forma canônica

$$\mathbf{M}^{-1}\mathbf{A}\mathbf{M} = \begin{bmatrix} 3 & -1 & -3 \\ -2 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 3 \\ -4 & 4 & 4 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 3 & -1 & -3 \\ -2 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

**4.7. Forma canônica real de Jordan.** O corpo dos complexos foi fundamental para garantir que temos autovalores suficientes para que o conjunto de subespaços generalizados gere o espaço todo. Caso isso aconteça no corpo dos reais, então também obtemos a forma canônica de Jordan.

**TEOREMA 4.9.** *Seja  $\mathbb{T} \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  de dimensão finita  $n = \dim(X) \in \mathbb{N}$  sobre  $\mathbb{R}$ . Sejam  $\lambda_1, \dots, \lambda_m$  os autovalores de  $\mathbb{T}$  em  $\mathbb{R}$  e suponha que os seus autoespaços generalizados geram todo o espaço, i.e.*

$$X = G(\mathbb{T}, \lambda_1) \oplus \dots \oplus G(\mathbb{T}, \lambda_m).$$

*Então existe uma base  $\mathcal{B}$  de  $X$  em que  $[\mathbb{T}]_{\mathcal{B}}$  está na forma canônica de Jordan.*

**DEMONSTRAÇÃO.** É só repetir a demonstração do [Teorema 4.8](#) a partir do momento em obtivemos que  $X$  é soma direta dos autoespaços generalizados.  $\square$

A condição do espaço ser soma direta dos autoespaços generalizados pode ser obtida através de um polinômio anulador.

**TEOREMA 4.10.** *Seja  $\mathbb{T} \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  de dimensão finita  $n = \dim(X) \in \mathbb{N}$  sobre  $\mathbb{R}$ . Suponha que  $p = p(x)$  seja um polinômio anulador de  $\mathbb{T}$  e que todas as raízes de  $p$  sejam reais. Então existe uma base  $\mathcal{B}$  de  $X$  em que  $[\mathbb{T}]_{\mathcal{B}}$  está na forma canônica de Jordan.*

**DEMONSTRAÇÃO.** É só aplicar o [Teorema 4.6](#) para deduzir que os autoespaços generalizados geram o espaço e, em seguida, aplicar [Teorema 4.9](#).  $\square$

Caso haja autovalores complexos, podemos *complexificar* o espaço e o operador, obter a forma canônica de Jordan complexa e voltar para os reais com uma forma canônica ligeiramente diferente.

DEFINIÇÃO 4.6. *Seja  $T \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  sobre o corpo dos reais. A **complexificação** de  $T$  é o operador linear  $T_{\mathbb{C}} \in \mathcal{L}(X_{\mathbb{C}})$ , no espaço  $X_{\mathbb{C}}$  complexificado (veja [Exemplo 1.8](#)), definido por*

$$(T|_{\mathbb{C}})\mathbf{u} = T\mathbf{u}_r + iT\mathbf{u}_i, \quad \forall \mathbf{u} = \mathbf{u}_r + i\mathbf{u}_i \in X_{\mathbb{C}}, \mathbf{u}_r, \mathbf{u}_i \in X.$$

No corpo dos complexos, a complexificação  $T_{\mathbb{C}}$  sempre tem uma forma canônica de Jordan. As questões são (i) como passar de uma base de  $X_{\mathbb{C}}$  para uma base de  $X$  e (ii) o que acontece com a forma canônica nessa passagem.

Vamos ver o que acontece em cada bloco. Primeiramente, se  $\lambda \in \mathbb{R}$  é um autovalor real do operador complexificado  $T_{\mathbb{C}}$ , com autovetor  $\mathbf{u} = \mathbf{u}_r + i\mathbf{u}_i$ , temos, por um lado,

$$T_{\mathbb{C}}(\mathbf{u}_r + i\mathbf{u}_i) = T(\mathbf{u}_r) + iT(\mathbf{u}_i),$$

e, por outro,

$$T_{\mathbb{C}}(\mathbf{u}_r + i\mathbf{u}_i) = \lambda(\mathbf{u}_r + i\mathbf{u}_i),$$

de modo que

$$T(\mathbf{u}_r) = \lambda\mathbf{u}_r, \quad T(\mathbf{u}_i) = \lambda\mathbf{u}_i.$$

Ou seja, com  $\lambda$  real, tanto a parte real quanto a parte imaginária de um autovetor complexo do operador complexificado são autovetores reais do operador real original. Na base  $\{\mathbf{u}_r, \mathbf{u}_i\}$ , ou em qualquer outra base, do subespaço  $S = \text{span}\{\mathbf{u}_r, \mathbf{u}_i\}$  de  $X$  gerado por esses dois vetores, temos

$$[T|_S]_{\{\mathbf{u}_r, \mathbf{u}_i\}} = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}.$$

Agora, se  $\lambda = \alpha + i\beta \in \mathbb{C}$ ,  $\alpha, \beta \in \mathbb{R}$  é um autovalor complexo de  $T_{\mathbb{C}}$ , com autovetor  $\mathbf{u} = \mathbf{u}_r + i\mathbf{u}_i$ , então, por um lado,

$$T_{\mathbb{C}}(\mathbf{u}) = (\alpha + i\beta)(\mathbf{u}_r + i\mathbf{u}_i) = (\alpha\mathbf{u}_r - \beta\mathbf{u}_i) + i(\beta\mathbf{u}_r + \alpha\mathbf{u}_i),$$

e, por outro,

$$T(\mathbf{u}_r) = \alpha\mathbf{u}_r - \beta\mathbf{u}_i,$$

$$T(\mathbf{u}_i) = \beta\mathbf{u}_r + \alpha\mathbf{u}_i.$$

Ou seja, na base  $\{\mathbf{u}_r, \mathbf{u}_i\}$  do subespaço  $S = \text{span}\{\mathbf{u}_r, \mathbf{u}_i\}$  de  $X$  gerado por esses dois vetores, temos

$$[T|_S]_{\{\mathbf{u}_r, \mathbf{u}_i\}} = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}.$$

Se  $\mathbf{u}_r, \mathbf{u}_i$  fossem a base canônica  $\mathbf{e}_1, \mathbf{e}_2$ , esta seria uma rotação no sentido horário, de um ângulo dado por

$$\theta = \arcsen \left( \frac{\beta}{\sqrt{\alpha^2 + \beta^2}} \right).$$

Para obtermos uma rotação no sentido anti-horário, ou seja, no sentido trigonométrico, devemos inverter a ordem, obtendo

$$[T]_{S|\{\mathbf{u}_i, \mathbf{u}_r\}} = \begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix}.$$

Por último, se

$$\mathbf{u} \mapsto \mathbf{N}_{\mathbb{C}}\mathbf{u} \mapsto \cdots \mapsto \mathbf{N}_{\mathbb{C}}^{k-1}\mathbf{u} \mapsto \mathbf{0}$$

é uma cadeia do operador nilpotente  $\mathbf{N}_{\mathbb{C}} = (\mathbf{T}_{\mathbb{C}} - \lambda I)_{G(\mathbb{T}_{\mathbb{C}}, \lambda)}$  associado a um autovalor  $\lambda$ , real ou complexo, então, como  $\mathbf{N}_{\mathbb{C}}\mathbf{u} = \mathbf{N}\mathbf{u}_r + i\mathbf{N}\mathbf{u}_i$ , obtemos as cadeias

$$\mathbf{u}_r \mapsto \mathbf{N}\mathbf{u}_r \mapsto \cdots \mapsto \mathbf{N}^{k-1}\mathbf{u}_r \mapsto \mathbf{0},$$

e

$$\mathbf{u}_i \mapsto \mathbf{N}\mathbf{u}_i \mapsto \cdots \mapsto \mathbf{N}^{k-1}\mathbf{u}_i \mapsto \mathbf{0}.$$

Em termos da base

$$\theta_{\lambda} = \{\mathbf{N}^{k-1}\mathbf{u}_i, \mathbf{N}^{k-1}\mathbf{u}_r, \dots, \mathbf{N}\mathbf{u}_i, \mathbf{N}\mathbf{u}_r, \mathbf{u}_i, \mathbf{u}_r\},$$

o operador  $\mathbf{N} = (\mathbf{T} - \lambda I)|_{\text{span}(\theta_{\lambda})}$  toma a forma

$$[\mathbf{N}]_{\theta_{\lambda}} = \begin{bmatrix} 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ & \vdots & \ddots & \ddots & & \vdots & \\ & \vdots & \ddots & \ddots & & 1 & 0 \\ & & & & & 0 & 1 \\ 0 & 0 & \cdots & \cdots & & 0 & 0 \\ 0 & 0 & \cdots & \cdots & & 0 & 0 \end{bmatrix},$$

que também pode ser escrito mais sucintamente na forma

$$[\mathbf{N}]_{\theta_{\lambda}} = \begin{bmatrix} \mathbf{0}_2 & \mathbf{I}_2 & & \\ & & \ddots & \\ & & & \mathbf{I}_2 \\ & & & & \mathbf{0}_2 \end{bmatrix},$$

onde  $\mathbf{I}_2$  é a matriz identidade em  $\mathbb{R}^2$  e  $\mathbf{0}_2$  é a matriz nula em  $\mathbb{R}^2$ . Com isso, obtemos, para  $\mathbb{T} - \lambda \mathbf{I}$  restrito a  $\text{span}(\mathfrak{b}_\lambda)$ , o *bloco de Jordan complexo*

$$[\mathbf{N}]_{\beta_\lambda} = J_{\lambda,k} = \begin{bmatrix} \alpha & -\beta & 1 & 0 & \cdots & 0 & 0 \\ \beta & \alpha & 0 & 1 & \cdots & 0 & 0 \\ & & \vdots & \ddots & \ddots & & \vdots \\ & & \vdots & \ddots & \ddots & 1 & 0 \\ & & & & & 0 & 1 \\ 0 & 0 & \cdots & \cdots & & \alpha & -\beta \\ 0 & 0 & & & & \beta & \alpha \end{bmatrix}.$$

Juntando esses diversos casos, obtemos a forma canônica real de Jordan.

**TEOREMA 4.11.** *Seja  $\mathbb{T} \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  de dimensão finita  $n = \dim(X) \in \mathbb{N}$  sobre  $\mathbb{R}$  e seja  $p = p(z)$  um polinômio anulador de  $\mathbb{T}$ , podendo ter raízes reais ou complexas. Uma raiz real corresponde a um autovalor de  $\mathbb{T}$ , que é também um autovalor do operador complexificado  $\mathbb{T}_{\mathbb{C}}$ , enquanto que uma raiz complexa é um autovalor apenas do operador complexificado  $\mathbb{T}_{\mathbb{C}}$ . Então existe uma base  $\mathfrak{b} = \mathfrak{b}_1 \cup \cdots \cup \mathfrak{b}_r$  de  $X$ , associada a uma soma direta  $X = G_1 \oplus \cdots \oplus G_r$ , com cada  $G_l$  invariante por  $\mathbb{T}$ , em que  $[\mathbb{T}]_{\mathfrak{b}}$  está na **forma canônica real de Jordan**, que tem a forma*

$$[\mathbb{T}]_{\mathfrak{b}} = \begin{bmatrix} \mathbf{J}_1 & & & \\ & \mathbf{J}_2 & & \\ & & \ddots & \\ & & & \mathbf{J}_r \end{bmatrix},$$

onde cada bloco  $\mathbf{J}_l = [\mathbb{T}|_{G_l}]_{\mathfrak{b}_l}$ ,  $l = 1, \dots, r$ ,  $r \in \mathbb{N}$ , tem uma das seguintes formas:

- (1)  $\mathbf{J}_l = \lambda_l \mathbf{I}$ , com  $\lambda_l \in \mathbb{R}$  um autovalor real de  $\mathbb{T}$ .
- (2)  $\mathbf{J}_l = \lambda_l \mathbf{I} + \mathbf{N}_l$ , com  $\lambda_l \in \mathbb{R}$  um autovalor real de  $\mathbb{T}$  e  $\mathbf{N}_l = (\delta_{i(j-1)})_{ij}$  é um bloco com 1 na superdiagonal e zero no resto, de modo que

$$\mathbf{J}_l = \lambda_l \mathbf{I} + \mathbf{N}_l = \begin{bmatrix} \lambda_l & 1 & & \\ & \lambda_l & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_l \end{bmatrix}.$$

- (3)  $\mathbf{J}_l$  é um bloco  $2 \times 2$  da forma

$$\mathbf{J}_l = \begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix},$$

com  $\alpha + i\beta$  um autovalor do operador complexificado  $\mathbb{T}_{\mathbb{C}}$ .

(4)  $\mathbf{J}_l$  é um bloco  $2k \times 2k$ , da forma

$$J_l = \begin{bmatrix} \alpha & -\beta & 1 & 0 & \cdots & 0 & 0 \\ \beta & \alpha & 0 & 1 & \cdots & 0 & 0 \\ & & \vdots & \ddots & \ddots & & \vdots \\ & & \vdots & \ddots & \ddots & 1 & 0 \\ & & & & & 0 & 1 \\ 0 & 0 & \cdots & \cdots & & \alpha & -\beta \\ 0 & 0 & \cdots & \cdots & & \beta & \alpha \end{bmatrix},$$

com  $\alpha + i\beta$  um autovalor do operador complexificado  $\mathbb{T}_{\mathbb{C}}$ .

## 5. Aplicações

**5.1. Sistemas de equações diferenciais lineares homogêneas e a exponencial de matrizes.** Um sistema homogêneo de  $n$  equações diferenciais lineares com coeficientes constantes tem a forma

$$\begin{cases} \frac{dx_1}{dt} = a_{11}x_1 + \dots + a_{1n}x_n, \\ \frac{dx_2}{dt} = a_{21}x_1 + \dots + a_{2n}x_n, \\ \vdots \\ \frac{dx_n}{dt} = a_{n1}x_1 + \dots + a_{nn}x_n, \end{cases} \quad (5.1)$$

com  $a_{ij} \in \mathbb{R}$ ,  $i, j = 1, \dots, n$ ,  $n \in \mathbb{N}$ . Podemos escrever o sistema (5.1) na forma vetorial

$$\frac{d}{dt} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

ou, de maneira mais sucinta,

$$\frac{d\mathbf{u}}{dt} = \mathbf{A}\mathbf{u}, \quad (5.2)$$

onde

$$\mathbf{u} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}.$$

No caso unidimensional, a equação toma a forma

$$\frac{dx}{dt} = ax,$$

para uma constante  $a$ , e cuja solução geral pode ser escrita como

$$x(t) = e^{at}x_0, \quad t \in \mathbb{R}.$$

Ficamos, naturalmente, tentados a escrever a solução geral de (5.2) como

$$\mathbf{u}(t) = e^{t\mathbf{A}}\mathbf{u}_0, \quad t \in \mathbb{R}, \quad (5.3)$$

onde trocamos a ordem para  $t\mathbf{A}$  visto que essa é a convenção no caso de multiplicação de uma matriz  $\mathbf{A}$  por um escalar  $t$ . A questão é como definir  $e^{t\mathbf{A}}$  em (5.3). Esperamos que seja uma matriz, com as mesmas dimensões de  $\mathbf{A}$ , da mesma forma que qualquer potência de  $\mathbf{A}$  é também uma matriz  $n \times n$ .

De fato, uma das maneiras de se escrever a exponencial  $e^a$  de um número real  $a$  é via série de potências

$$e^a = 1 + a + \frac{a^2}{2} + \cdots + \frac{a^j}{j!} + \cdots = \sum_{j=0}^{\infty} \frac{a^j}{j!}.$$

Há várias outras fórmulas possíveis de se escrever a função exponencial, como

$$\begin{aligned} e^a &= \sum_{n=0}^{\infty} \frac{a^n}{n!} \\ e^a &= \lim_{n \rightarrow \infty} \left(1 + \frac{a}{n}\right)^n; \\ e^a &= \lim_{n \rightarrow \infty} \left(1 - \frac{a}{n}\right)^{-n}; \\ e^a &= \frac{1}{2\pi i} \int_{\gamma} \frac{e^{\zeta}}{\zeta - a} d\zeta; \end{aligned}$$

onde  $\gamma$  é uma curva fechada simples, no plano complexo, envolvendo  $a$ .

**OBSERVAÇÃO 5.1.** Por exemplo, sendo  $\mathbf{A}$  um operador positivo e  $\gamma$  uma curva fechada no semiplano complexo  $\operatorname{Re}(z) > 0$  envolvendo o espectro de  $\mathbf{A}$ , conceitos que iremos ver mais pra frente, a exponencial do operador  $t\mathbf{A}$  pode ser definida, para  $t \geq 0$ , através da fórmula de Cauchy

$$e^{t\mathbf{A}} = \frac{1}{2\pi i} \int_{\gamma} e^{\zeta} (\zeta - t\mathbf{A})^{-1} d\zeta.$$

Variáveis complexas são de grande importância no estudo de operadores lineares, facilitando, por exemplo, a demonstração de regularidade de soluções de equações diferenciais (parciais, inclusive, com os semigrupos analíticos) e em análise de perturbação do espectro, sensibilidade e robustez de sistemas.

No caso de um operador linear  $A$ , podemos definir  $e^A$  de qualquer uma das maneiras acima, mas, para isso, é necessário considerarmos alguma métrica, para estabelecer as convergências (veja e.g. [Observação 1.12](#) e [Seção 5.1](#)).

Uma outra maneira, é usar a forma canônica de Jordan para definir a exponencial de uma maneira mais explícita, conforme fazemos a seguir. A vantagem de usar a forma canônica é que as fórmulas acima se reduzem a limites de função reais, associada a cada autovalor de  $T$ , combinadas com somas finitas de operadores nilpotentes, evitando trabalharmos com a convergência no espaço de matrizes.

**TEOREMA 5.1.** *Seja  $T = \lambda I + N$  um operador linear em um espaço vetorial  $X$ , com  $\lambda \in \mathbb{K}$  e  $N$  nilpotente de índice  $k$ . Então*

$$(\lambda I + N)^j = \sum_{i=0}^j \binom{j}{i} \lambda^i N^{j-i} = \sum_{i=\min\{0, j-k+1\}}^j \frac{j!}{j!(j-i)!} \lambda^i N^{j-i},$$

para todo  $j = 0, 1, \dots$ .

**DEMONSTRAÇÃO.** Como  $I$  e  $N$  comutam, podemos expandir a potência como no binômio de Newton, nos dando

$$(\lambda I + N)^j = \sum_{i=0}^j \binom{j}{i} \lambda^i N^{j-i}.$$

Agora, como  $N$  é nilpotente de índice  $k$ , segue que  $N^{j-i} = 0$ , para  $j - i \geq k$ , ou seja, só precisamos incluir no somatório os índices  $i$  tais que  $j - i \leq k - 1$ , ou seja,  $i \geq j - k + 1$ . Como  $i$  também deve ser não-nulo, chegamos ao somatório com índice  $i$  satisfazendo  $\min\{0, j - k + 1\} \leq i \leq j$ , provando o resultado.  $\square$

**TEOREMA 5.2.** *Seja  $T = \lambda I + N$  um operador linear em um espaço vetorial  $X$ , com  $\lambda \in \mathbb{K}$  e  $N$  nilpotente de índice  $k$ . Então*

$$\sum_{j=0}^m \frac{1}{j!} (\lambda I + N)^j = \left( \sum_{i=0}^m \frac{1}{i!} \lambda^i \right) \left( \sum_{l=0}^{k-1} \frac{1}{l!} N^l \right). \quad (5.4)$$

DEMONSTRAÇÃO. Uando o **Teorema 5.1**, obtemos

$$\begin{aligned} \sum_{j=0}^m \frac{1}{j!} (\lambda \mathbf{I} + \mathbf{N})^j &= \sum_{j=0}^m \frac{1}{j!} \sum_{i=\min\{0, j-k+1\}}^j \frac{j!}{i!(j-i)!} \lambda^i \mathbf{N}^{j-i} \\ &= \sum_{j=0}^m \sum_{i=\min\{0, j-k+1\}}^j \frac{1}{i!(j-i)!} \lambda^i \mathbf{N}^{j-i}. \end{aligned}$$

Trocando a ordem do somatório, a condição  $i \geq j - k + 1$  se transpõe para a condição  $j \leq i + k - 1$ , nos dando

$$\sum_{j=0}^m \frac{1}{j!} (\lambda \mathbf{I} + \mathbf{N})^j = \sum_{i=0}^m \sum_{j=i}^{i+k-1} \frac{1}{i!(j-i)!} \lambda^i \mathbf{N}^{j-i}.$$

Mudando o somatório interno para  $l = j - i$ , desacoplamos os somatórios, obtendo

$$\sum_{j=0}^m \frac{1}{j!} (\lambda \mathbf{I} + \mathbf{N})^j = \sum_{i=0}^m \sum_{l=0}^{k-1} \frac{1}{i!l!} \lambda^i \mathbf{N}^l = \left( \sum_{i=0}^m \frac{1}{i!} \lambda^i \right) \left( \sum_{l=0}^{k-1} \frac{1}{l!} \mathbf{N}^l \right).$$

□

Fazendo  $m \rightarrow \infty$  na fórmula (5.4), observamos que o primeiro somatório do lado direito é um somatório escalar que converge, em  $\mathbb{R}$ , para a exponencial de  $\lambda$ , enquanto que o segundo somatório é independente de  $m$ . Dessa forma, definimos a exponencial do bloco de Jordan através do limite do lado direito.

**DEFINIÇÃO 5.1** (Exponencial de um bloco de Jordan). *Seja  $\mathbf{T} = \lambda \mathbf{I} + \mathbf{N}$  um operador linear em um espaço vetorial  $X$ , com  $\lambda \in \mathbb{K}$  e  $\mathbf{N}$  nilpotente de índice  $k$ . Então definimos a exponencial de  $\mathbf{T}$  pela fórmula*

$$e^{\mathbf{T}} = \sum_{j=0}^{\infty} \frac{1}{j!} (\lambda \mathbf{I} + \mathbf{N})^j = e^{\lambda} \left( \mathbf{I} + \mathbf{N} + \dots + \frac{1}{(k-1)!} \mathbf{N}^{k-1} \right).$$

**OBSERVAÇÃO 5.2.** A **Definição 5.1** não trata, exatamente, de um bloco de Jordan, conforme anunciado no título da definição, mas é um bloco de Jordan em espírito. Ou seja, em uma base  $\mathcal{B}$  apropriada,  $[\mathbf{T}]_{\mathcal{B}} = \lambda \mathbf{I} + \mathbf{N}$  está na forma canônica de um bloco de Jordan.

**OBSERVAÇÃO 5.3.** Quando falarmos de espaços vetoriais normados e operadores lineares contínuos, veremos que podemos definir  $e^{\mathbf{T}} = \sum_{j=0}^{\infty} \mathbf{T}^j / j!$ , para um operador linear  $\mathbf{T} \in \mathcal{L}(X)$  qualquer em um espaço de dimensão finita  $X$ , ou para um operador

linear *limitado*, ou *contínuo*, em um espaço vetorial qualquer  $X$ . Nesse caso, segue, naturalmente, que  $e^T$  coincide com a fórmula dada no [Definição 5.1](#), para  $T = \lambda I + N$ .

**COROLÁRIO 5.1** (Exponencial de matriz similar a um bloco de Jordan). *Seja  $A \in \mathbb{K}^{k \times k}$ ,  $k \in \mathbb{N}$ , uma matriz quadrada. Suponha que exista uma matriz invertível  $M \in \mathbb{K}^{k \times k}$  tal que  $B = M^{-1}AM = \lambda I + N$ , onde  $\lambda \in \mathbb{K}$  e  $N = (\delta_{i(j-1)})$  é nilpotente na forma canônica, de índice  $k$ . Então a exponencial de  $A$  é dada pela fórmula*

$$e^A = Me^BM^{-1} = M \left( e^{\lambda(I + N + \dots + \frac{1}{(k-1)!}N^{k-1})} \right) M$$

$$= e^{\lambda M} \begin{bmatrix} 1 & 1 & \frac{1}{2} & \frac{1}{6} & \dots & \frac{1}{(k-1)!} \\ 0 & 1 & 1 & \frac{1}{2} & \dots & \frac{1}{(k-2)!} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \frac{1}{2} \\ \vdots & & & \ddots & \ddots & 1 \\ 0 & \dots & \dots & \dots & 0 & 1 \end{bmatrix}.$$

O resultado anterior diz respeito a um bloco. Mas, em geral, temos vários blocos. Mas como são independentes (i.e. agindo em subespaços invariante independentes), as potências não atrapalham a estrutura de blocos e a estrutura segue até o limite, com a exponencial. Mais precisamente, temos o seguinte resultado.

**TEOREMA 5.3.** *Seja  $A \in \mathbb{K}^{n \times n}$  e suponha que exista  $M \in \mathbb{K}^{n \times n}$  tal que  $B = M^{-1}AM$  esteja na forma canônica de Jordan*

$$B = M^{-1}AM = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_r \end{bmatrix},$$

onde cada bloco  $J_l$ ,  $l = 1, \dots, r$ , é da forma  $\lambda_l I + N_l$ , onde  $\lambda_l \in \mathbb{K}$  é um autovalor de  $A$  e  $N_l = (\delta_{i(j-1)})_{ij}$  é um bloco com 1 na superdiagonal e zero no resto. Então

$$e^A = Me^BM^{-1} = M \begin{bmatrix} e^{J_1} & & & \\ & e^{J_2} & & \\ & & \ddots & \\ & & & e^{J_r} \end{bmatrix} M^{-1}.$$

Mais coisas depois...

## 6. Exercícios

## Exercícios

- 6.1. Formule uma versão do Teorema 3.2 para obter uma representação do operador na forma triangular inferior.
- 6.2. Formule e demonstre uma versão do Teorema 3.3 para a existência de uma forma triangular inferior.
- 6.3. Formule e demonstre uma versão do Teorema 3.4 para a diagonal da forma triangular inferior de um operador em um espaço vetorial complexo não trivial de dimensão finita.
- 6.4. Encontre o maior subespaço de Krylov  $\mathcal{K}_k(\mathbf{A}, \mathbf{u}) = \text{span}\{\mathbf{u}, \mathbf{A}\mathbf{u}, \dots, \mathbf{A}^{k-1}\mathbf{u}\}$  associado ao vetor  $\mathbf{u} = (1, 1, 1, 1)$  e à matriz

$$\mathbf{A} = \begin{bmatrix} 4 & -1 & 0 & 0 \\ 10 & -2 & 0 & -1 \\ -1 & 0 & 2 & 1 \\ 2 & -1 & 0 & 2 \end{bmatrix}.$$

- 6.5. Considere a matriz

$$\mathbf{A} = \begin{bmatrix} -4 & 21 & -16 \\ -5 & 18 & -12 \\ -4 & 12 & -7 \end{bmatrix}$$

e o vetor  $\mathbf{u} = (5, 0, 2)$ . Mostre que  $\{\mathbf{u}, \mathbf{A}\mathbf{u}, \mathbf{A}^2\mathbf{u}\}$  é linearmente dependente e ache um polinômio  $p(z) = az^2 + bz + c$  tal que  $a\mathbf{A}^2\mathbf{u} + b\mathbf{A}\mathbf{u} + c\mathbf{u} = \mathbf{0}$ . Use esse polinômio para encontrar um dos autovalores de  $\mathbf{A}$  e um autovetor associado.

- 6.6. Considere, novamente, a matriz do Exercício 6.5,

$$\mathbf{A} = \begin{bmatrix} -4 & 21 & -16 \\ -5 & 18 & -12 \\ -4 & 12 & -7 \end{bmatrix}$$

mas agora com o vetor  $\mathbf{u} = (1, 0, 0)$ . Mostre que  $\{\mathbf{u}, \mathbf{A}\mathbf{u}, \mathbf{A}^2\mathbf{u}, \mathbf{A}^3\mathbf{u}\}$  é linearmente independente e ache um polinômio  $p(z) = az^3 + bz^2 + cz + d$  tal que  $a\mathbf{A}^3\mathbf{u} + b\mathbf{A}^2\mathbf{u} + c\mathbf{A}\mathbf{u} + d\mathbf{u} = \mathbf{0}$  e use esse polinômio para encontrar o outro autovalor de  $\mathbf{A}$  (No Exercício 6.9, pedimos para encontrar os autoespaços generalizados e, em particular, deduzir que esses são os dois únicos autovalores de  $\mathbf{A}$ ).

- 6.7.** No [Teorema 4.1](#), se  $\mathbf{u} \in G(\mathbb{T}, \lambda)$  é um autovetor generalizado com índice  $k$ , mostre que  $\mathbf{v} = \mathbb{T}\mathbf{u}$  também é um autovalor generalizado, associado ao autovalor  $\lambda$ , com índice  $k$ , caso  $\lambda \neq 0$ , e índice  $k - 1$ , caso  $\lambda = 0$ .
- 6.8.** Seja  $\mathbb{T} \in \mathcal{L}(X)$  um operador linear em um espaço vetorial  $X$  de dimensão finita sobre um corpo  $\mathbb{K}$ . Suponha que  $\lambda_1, \dots, \lambda_m$  sejam autovalores “completos” de  $\mathbb{T}$ , no sentido de que

$$X = G(\mathbb{T}, \lambda_1) \oplus \cdots \oplus G(\mathbb{T}, \lambda_m).$$

Já vimos que cada  $G(\mathbb{T}, \lambda_j)$  é invariante por  $\mathbb{T}$  e que a restrição  $(\mathbb{T} - \lambda_j I)|_{G(\mathbb{T}, \lambda_j)}$  de  $\mathbb{T} - \lambda_j I$  a  $G(\mathbb{T}, \lambda_j)$  é nilpotente. Mostre, agora, que a restrição  $(\mathbb{T} - \lambda_j I)|_{G(\mathbb{T}, \lambda_i)}$  de  $\mathbb{T} - \lambda_j I$  ao autoespaço generalizado  $G(\mathbb{T}, \lambda_i)$  associado a  $\lambda_i \neq \lambda_j$  é invertível.

- 6.9.** Encontre a forma canônica de Jordan da seguinte matriz considerada no [Exercício 6.5](#) e no [Exercício 6.6](#),

$$\mathbf{A} = \begin{bmatrix} -4 & 21 & -16 \\ -5 & 18 & -12 \\ -4 & 12 & -7 \end{bmatrix}.$$

- 6.10.** Encontre a forma canônica de Jordan da matriz

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 \\ 2 & 4 & 1 \\ 0 & 0 & 2 \end{bmatrix}$$

- 6.11.** Encontre a forma canônica de Jordan da matriz

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ -7 & 3 & 3 & 3 \\ -1 & 0 & 2 & 1 \\ 4 & -1 & 0 & -1 \end{bmatrix}$$

- 6.12.** Encontre a forma canônica de Jordan da matriz

$$\mathbf{A} = \begin{bmatrix} 5 & -1 & 0 & -2 \\ 5 & 0 & 0 & -3 \\ 3 & -1 & 1 & -1 \\ 4 & -1 & 0 & -1 \end{bmatrix}$$

- 6.13.** Encontre a forma canônica de Jordan da matriz

$$\mathbf{A} = \begin{bmatrix} 4 & -1 & 0 & 0 \\ 10 & -2 & 0 & -1 \\ -1 & 0 & 2 & 1 \\ 2 & -1 & 0 & 2 \end{bmatrix}$$

- 6.14.** Verifique que o operador complexificado dado na [Definição 4.6](#) é, de fato, um operador linear sobre o espaço complexificado.



## CAPÍTULO 7

### Norma e Produto Interno

#### 1. Norma

A noção de continuidade é um dos conceitos fundamentais em análise e em matemática em geral. Ela depende, muitas vezes, de alguma noção de distância. Mais geralmente, podemos falar de continuidade a partir de conjuntos que denominamos de *abertos* e que satisfazem certas propriedades que encompassam alguma noção de proximidade, nos dando o que chamamos de *topologia* de um espaço. Felizmente, em espaços vetoriais de dimensão finita, não precisamos dessa generalidade toda. Podemos nos restringir ao conceito de *norma*, expressando o “comprimento” de um vetor, e com base nele podemos falar de distância, topologia e continuidade.

**1.1. Definição e exemplos.** Espaços métricos são conjuntos munidos de uma medida de “distância”  $d(x, y)$  entre dois pontos  $x$  e  $y$  quaisquer, satisfazendo propriedades adequadas.

**DEFINIÇÃO 1.1.** Um **espaço métrico** é um conjunto  $X$  munido de uma função distância  $d : X \times X \rightarrow \mathbb{R}$  que é positiva definida, simétrica e satisfaz a desigualdade triangular, i.e.

- (i)  $d(x, y) > 0$  para  $x \neq y$ ,  $x, y \in X$  (distância entre pontos distintos é sempre positiva);
- (ii)  $d(x, x) = 0$ , para todo  $x \in X$  (um ponto não pode estar distante de si mesmo);
- (iii)  $d(x, y) = d(y, x)$  para quaisquer  $x, y \in X$  (a distância de  $x$  a  $y$  é igual à distância de  $y$  a  $x$ ); e
- (iv)  $d(x, z) \leq d(x, y) + d(y, z)$ , para quaisquer  $x, y, z \in X$  (a distância mede o “caminho” mais curto entre dois pontos - veja **Observação 1.4**)

Ao combinarmos a noção de métrica com a estrutura de um espaço vetorial, esperamos que uma métrica tenha propriedades compatíveis com essa estrutura, como, em particular, que as operações de adição e multiplicação por escalar sejam contínuas em relação a essa métrica.

Além disso, é natural esperarmos que essa métrica seja *homogênea*, como é a distância usual no plano, por exemplo, onde o comprimento de um segmento de reta

independe da sua posição no plano. Ou seja, em geral, esperamos que a distância entre dois vetores  $\mathbf{u}$  e  $\mathbf{v}$  seja a mesma que a distância entre  $\mathbf{u} + \mathbf{w}$  e  $\mathbf{v} + \mathbf{w}$ , para qualquer  $\mathbf{w}$ . Ou, por outra, que a distância entre  $\mathbf{u}$  e  $\mathbf{v}$  seja a mesma que entre  $\mathbf{0}$  e  $\mathbf{u} - \mathbf{v}$ . A distância entre um vetor  $\mathbf{u}$  e o vetor nulo  $\mathbf{0}$  mede o *comprimento* de um vetor. E ao multiplicarmos um vetor por um escalar, esperamos aumentar o seu comprimento pelo mesmo fator, ou, na verdade, pelo módulo desse escalar, com o sinal apenas mudando o vetor de sentido, mas sem alterar o seu comprimento.

Essa ideia nos dá a noção de *norma*, que definimos a seguir.

DEFINIÇÃO 1.2. Uma **norma** em um espaço vetorial  $X$  sobre um corpo  $\mathbb{K}$  real ou complexo é uma função  $\|\cdot\| : X \rightarrow \mathbb{R}$  positiva definida, homogênea e satisfazendo a desigualdade triangular, ou seja

- (i) (positiva definida)  $\|\mathbf{u}\| \geq 0$ , para todo  $\mathbf{u} \in X$  com  $\|\mathbf{u}\| = 0$  se, e somente se,  $\mathbf{u} = \mathbf{0}$ ;
- (ii) (homogênea)  $\|\lambda\mathbf{u}\| = |\lambda|\|\mathbf{u}\|$ , para todo  $\mathbf{u} \in X$ ;
- (iii) (desigualdade triangular)  $\|\mathbf{u} + \mathbf{w}\| \leq \|\mathbf{u}\| + \|\mathbf{w}\|$ .

Um espaço vetorial munido de uma norma é chamado de **espaço vetorial normado**.

OBSERVAÇÃO 1.1. A métrica associada a uma norma  $\|\cdot\|$  em um espaço vetorial  $X$  é dada por

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|.$$

A simetria da métrica sai da homogeneidade da norma com  $\lambda = -1$ , visto que

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \|\mathbf{v} - \mathbf{u}\| = d(\mathbf{v}, \mathbf{u}).$$

A desigualdade triangular da métrica sai diretamente da desigualdade triangular da norma, visto que

$$\begin{aligned} d(\mathbf{u}, \mathbf{v}) &= \|\mathbf{u} - \mathbf{v}\| = \|\mathbf{u} - \mathbf{w} + \mathbf{w} - \mathbf{v}\| \\ &\leq \|\mathbf{u} - \mathbf{w}\| + \|\mathbf{w} - \mathbf{v}\| \leq d(\mathbf{u}, \mathbf{w}) + d(\mathbf{w}, \mathbf{v}). \end{aligned}$$

OBSERVAÇÃO 1.2 (Continuidade da adição). A adição é uma função  $+: X \times X \rightarrow X$ . A continuidade, em um ponto  $(\mathbf{u}, \mathbf{v}) \in X \times X$ , da soma  $\mathbf{u} + \mathbf{v}$  é obtida comparando-se a distância entre a soma  $\tilde{\mathbf{u}} + \tilde{\mathbf{v}}$  de uma perturbação  $\tilde{\mathbf{u}} \approx \mathbf{u}$  de  $\mathbf{u}$  e uma perturbação  $\tilde{\mathbf{v}} \approx \mathbf{v}$  e a soma  $\mathbf{u} + \mathbf{v}$  de  $\mathbf{u}$  e  $\mathbf{v}$ . Essa continuidade segue da desigualdade

$$d(\tilde{\mathbf{u}} + \tilde{\mathbf{v}}, \mathbf{u} + \mathbf{v}) = \|\tilde{\mathbf{u}} + \tilde{\mathbf{v}} - \mathbf{u} - \mathbf{v}\| \leq \|\tilde{\mathbf{u}} - \mathbf{u}\| + \|\tilde{\mathbf{v}} - \mathbf{v}\| = d(\tilde{\mathbf{u}}, \mathbf{u}) + d(\tilde{\mathbf{v}}, \mathbf{v}).$$

Assim,  $\tilde{\mathbf{u}} + \tilde{\mathbf{v}} \rightarrow \mathbf{u} + \mathbf{v}$ , quando  $\tilde{\mathbf{u}} \rightarrow \mathbf{u}$  e  $\tilde{\mathbf{v}} \rightarrow \mathbf{v}$ . Quanto mais próximos  $\tilde{\mathbf{u}} \approx \mathbf{u}$  e  $\tilde{\mathbf{v}} \approx \mathbf{v}$ , mais próximo  $\tilde{\mathbf{u}} + \tilde{\mathbf{v}}$  está de  $\mathbf{u} + \mathbf{v}$ .

OBSERVAÇÃO 1.3 (Continuidade da multiplicação por escalar). De maneira semelhante,

$$\begin{aligned}
 d(\tilde{\lambda}\tilde{\mathbf{u}}, \lambda\mathbf{u}) &= \|\tilde{\lambda}\tilde{\mathbf{u}} - \lambda\mathbf{u}\| \\
 &= \|\tilde{\lambda}\tilde{\mathbf{u}} - \lambda\tilde{\mathbf{u}} + \lambda\tilde{\mathbf{u}} - \lambda\mathbf{u}\| \\
 &\leq \|\tilde{\lambda}\tilde{\mathbf{u}} - \lambda\tilde{\mathbf{u}}\| + \|\lambda\tilde{\mathbf{u}} - \lambda\mathbf{u}\| \\
 &= \|(\tilde{\lambda} - \lambda)\tilde{\mathbf{u}}\| + \|\lambda(\tilde{\mathbf{u}} - \mathbf{u})\| \\
 &= |\tilde{\lambda} - \lambda|\|\tilde{\mathbf{u}}\| + |\lambda|\|\tilde{\mathbf{u}} - \mathbf{u}\| \\
 &= |\tilde{\lambda} - \lambda|\|\tilde{\mathbf{u}}\| + |\lambda|d(\tilde{\mathbf{u}}, \mathbf{u}).
 \end{aligned}$$

Assim, quando  $\tilde{\mathbf{u}} \rightarrow \mathbf{u}$  em  $X$  e  $\tilde{\lambda} \rightarrow \lambda$  em  $\mathbb{K}$ , temos  $\tilde{\lambda}\tilde{\mathbf{u}} \rightarrow \lambda\mathbf{u}$ , nos dando a continuidade da multiplicação por escalar.

OBSERVAÇÃO 1.4 (Caminho mais curto). Dada uma métrica  $d(\cdot, \cdot)$  em um conjunto  $X$ , a desigualdade triangular garante que a distância é a menor possível entre dois pontos, no sentido de que

$$d(x, y) = \min_{z \in X} (d(x, z) + d(z, y)),$$

para todo  $x, y \in X$ . De fato, a desigualdade triangular garante que

$$d(x, y) \leq d(x, z) + d(z, y), \quad \forall z \in X.$$

Agora, escolhendo  $z = x$  (idem para  $z = y$ ), vemos que

$$d(x, z) + d(z, y) = d(x, x) + d(x, y) = d(x, y),$$

visto que  $d(x, x) = 0$ . Portanto, o mínimo é alcançado e é igual a  $d(x, y)$ , estabelecendo a igualdade acima. Podemos interpretar o resultado como dizendo que a distância mede o “caminho” mais curto entre dois pontos. Em um espaço vetorial, o conceito de caminho faz sentido, pois podemos considerar o segmento de reta  $(1-\theta)x + \theta y$ ,  $0 \leq \theta \leq 1$ . No caso geral de um espaço métrico, não temos essas operações de adição e multiplicação por escalar. No entanto, esse é o espírito, quando dizemos que, se incluirmos qualquer ponto  $z$  na medição, ou se tentarmos medir a distância entre dois pontos a partir da distância deles a um terceiro ponto, não conseguiremos uma medida mais curta.

EXEMPLO 1.1. Em  $\mathbb{R}^n$  ou  $\mathbb{C}^n$ , temos a norma  $p$

$$\|\mathbf{u}\|_p = \left( \sum_{j=1}^n |x_j|^p \right)^{1/p},$$

para todo vetor  $\mathbf{u} = (x_1, \dots, x_n) \in \mathbb{K}^n$ ,  $\mathbb{K}$  real ou complexo, e onde  $p$  é qualquer real no intervalo  $1 \leq p < \infty$ . O caso  $p = \infty$  nos dá, como limite  $p \rightarrow \infty$ ,

$$\|\mathbf{u}\|_p = \max\{|x_1|, \dots, |x_n|\}.$$

A norma  $p$  com  $1 < p < \infty$  tem a propriedade de ser estritamente convexa, ou seja, se  $\|\mathbf{u}\| = \|\mathbf{v}\| = r$ , com  $\mathbf{u} \neq \mathbf{v}$ , então

$$\|(1 - \theta)\mathbf{u} + \theta\mathbf{v}\| < r,$$

para  $0 < \theta < 1$ , ou seja, o interior do segmento de reta entre dois pontos de uma esfera está contido inteiramente no interior da esfera. Isso não acontece com  $p = 1$  ou  $p = \infty$ . Mais ainda, para  $1 < p < \infty$ , o ponto médio está uniformemente bem afastado da esfera, i.e. para todo  $\varepsilon > 0$ , existe  $\delta > 0$  tal que se  $\|\mathbf{u}\|_p = \|\mathbf{v}\|_p = 1$  e  $\|\mathbf{u} - \mathbf{v}\| \geq \varepsilon$ , então

$$\left\| \frac{x + y}{2} \right\| \leq 1 - \delta.$$

Com essa propriedade, dizemos que a norma é *uniformemente convexa*.

EXEMPLO 1.2 (Norma em  $\ell^p$ ). Podemos estender a norma  $p$  para seqüências reais ou complexas, desde que o somatório seja finito. Isso é obtido restringindo as seqüências ao espaço  $\ell^p$  (veja [Exemplo 1.10](#)), mas aqui vamos assumir  $1 \leq p < \infty$ . Temos

$$\ell^p = \{x = (x_n)_{n \in \mathbb{N}}; \sum_{n \in \mathbb{N}} |x_n|^p < \infty\}.$$

É fácil verificar que, para  $1 \leq p < \infty$ , temos

$$\|\mathbf{u}\|_{\ell^p} = \left( \sum_{n \in \mathbb{N}} |x_n|^p \right)^{1/p}$$

satisfazendo as hipóteses de norma. Para  $p = \infty$ , temos seqüências limitadas

$$\ell^\infty = \{x = (x_n)_{n \in \mathbb{N}}; \sup_{n \in \mathbb{N}} |x_n| < \infty\},$$

com a norma do supremo

$$\|\mathbf{u}\|_{\ell^\infty} = \sup_{n \in \mathbb{N}} |x_n|.$$

EXEMPLO 1.3. No espaço das funções reais contínuas  $X = \mathcal{C}(I)$  em um intervalo  $I \subset \mathbb{R}$  qualquer, por exemplo, podemos definir a norma  $p$  por

$$\|f\|_p = \left( \int_I |f(x)|^p dx \right)^{1/p},$$

para  $1 \leq p < \infty$  e

$$\|f\|_\infty = \sup_{x \in I} |f(x)|.$$

No caso em que  $I$  é fechado e limitado, o supremo se torna um máximo.

Vale ressaltar que a própria norma é uma função contínua de  $X$  em  $\mathbb{R}$ :

**TEOREMA 1.1** (Continuidade da norma). *Dado um espaço vetorial  $X$  real ou complexo munido de uma norma  $\|\cdot\|$ , vale a desigualdade*

$$|\|\mathbf{u}\| - \|\mathbf{v}\|| \leq \|\mathbf{u} - \mathbf{v}\|, \quad \forall \mathbf{u}, \mathbf{v} \in X.$$

**DEMONSTRAÇÃO.** Pela desigualdade triangular,

$$\|\mathbf{u}\| \leq \|\mathbf{u} - \mathbf{v}\| + \|\mathbf{v}\|,$$

de modo que

$$\|\mathbf{u}\| - \|\mathbf{v}\| \leq \|\mathbf{u} - \mathbf{v}\|.$$

Da mesma forma,

$$\|\mathbf{v}\| - \|\mathbf{u}\| \leq \|\mathbf{v} - \mathbf{u}\|.$$

Pela homogeneidade, com  $\lambda = -1$ , temos

$$\|\mathbf{v} - \mathbf{u}\| = \|\mathbf{u} - \mathbf{v}\|,$$

de modo que

$$-\|\mathbf{u} - \mathbf{v}\| \leq \|\mathbf{u}\| - \|\mathbf{v}\| \leq \|\mathbf{u} - \mathbf{v}\|,$$

ou seja,

$$|\|\mathbf{u}\| - \|\mathbf{v}\|| \leq \|\mathbf{u} - \mathbf{v}\|,$$

provando a desigualdade desejada.  $\square$

**1.2. Topologia de um espaço vetorial normado.** Com a noção de distância dada pela métrica, podemos falar de *convergência* de sequências. Também podemos falar de subconjuntos *abertos* e *fechados* e de *topologia*.

**DEFINIÇÃO 1.3** (Convergência de sequências). *Seja  $X$  um espaço vetorial munido de uma norma  $\|\cdot\|$ . Dizemos que uma sequência  $(\mathbf{u}_k)_{k \in \mathbb{N}}$  em  $X$  converge para um vetor  $\mathbf{u} \in X$  quando*

$$\lim_{k \rightarrow \infty} \|\mathbf{u}_k - \mathbf{u}\| = 0.$$

Nesse caso, escrevemos

$$\lim_{k \rightarrow \infty} \mathbf{u}_k = \mathbf{u}$$

ou

$$\mathbf{u}_k \rightarrow \mathbf{u}, \quad \text{quando } k \rightarrow \infty.$$

O limite é único, como demonstra o próximo resultado.

**TEOREMA 1.2.** *Seja  $X$  um espaço vetorial normado real ou complexo e suponha que  $(\mathbf{u}_k)_{k \in \mathbb{N}}$  seja uma sequência convergente. Então o limite da sequência é único.*

**DEMONSTRAÇÃO.** De fato, se  $\mathbf{u}_k \rightarrow \mathbf{u}$  e  $\mathbf{u}_k \rightarrow \mathbf{v}$ , então, pela desigualdade triangular,

$$\|\mathbf{u} - \mathbf{v}\| \leq \|\mathbf{u} - \mathbf{u}_k\| + \|\mathbf{u}_k - \mathbf{v}\| \rightarrow 0,$$

quando  $k \rightarrow \infty$ . Como a norma é positiva definida, temos, necessariamente,  $\mathbf{u} - \mathbf{v} = \mathbf{0}$ , ou seja,  $\mathbf{u} = \mathbf{v}$ , mostrando que o limite é único.  $\square$

Uma propriedade fundamental dos reais, associada à sua própria construção e à sua razão de existência, é a de ser *completo*, no sentido de que as *seqüências de Cauchy* reais, que são as seqüências reais  $\{r_n\}$  tais que  $|r_n - r_m| \rightarrow 0$ , quando  $n, m \rightarrow \infty$ , convergem para um número real  $r$ . Trabalhando em cada coordenada, essa mesma propriedade vale em  $\mathbb{R}^n$ . E vale, também, em  $\mathbb{C}^n$ , separando as partes real e imaginária de uma seqüência de Cauchy de vetores complexos. Finalmente, graças à representação de espaços vetoriais de dimensão finita como  $\mathbb{K}^n$ , em um corpo  $\mathbb{K}$ , e graças à equivalência das normas que veremos em seguida, deduz-se que espaços vetoriais complexos ou reais de dimensão finita são também completos.

**DEFINIÇÃO 1.4.** *Seja  $X$  um espaço vetorial real ou complexo munido de uma norma  $\|\cdot\|_X$ . Uma seqüência  $\{\mathbf{u}_k\}_{k \in \mathbb{N}}$  em  $X$  é dita uma **seqüência de Cauchy** em  $X$  quando, para todo  $\varepsilon > 0$ , existe  $K \in \mathbb{N}$  tal que*

$$\|\mathbf{u}_k - \mathbf{u}_l\|_X < \varepsilon, \quad \forall k, l \geq K,$$

ou seja,  $\|\mathbf{u}_k - \mathbf{u}_l\|_X \rightarrow 0$ , quando  $k, l \rightarrow \infty$ . Além disso, o espaço  $X$  é dito **completo** quando toda seqüência de Cauchy é convergente, i.e. dada uma seqüência de Cauchy  $\{\mathbf{u}_k\}_{k \in \mathbb{N}}$ , existe  $\mathbf{u} \in X$  tal que  $\mathbf{u} = \lim_{k \rightarrow \infty} \mathbf{u}_k$ .

A demonstração de completude depende da equivalência entre a norma de  $X$  e a norma de  $\mathbb{K}^n$  via isomorfismo entre  $X$  e  $\mathbb{K}^n$ , com  $n = \dim(X)$ . Isso só será visto no **Teorema 1.13**. De qualquer forma, anunciamos e demonstramos a completude de  $X$  a seguir.

**TEOREMA 1.3** (Completude de espaços vetoriais de dimensão finita). *Seja  $X$  um espaço vetorial real ou complexo de dimensão finita munido de uma norma  $\|\cdot\|_X$ . Então  $X$  é completo, i.e. toda seqüência de Cauchy em  $X$  é convergente.*

**DEMONSTRAÇÃO.** Seja  $\{\mathbf{u}_k\}_{k \in \mathbb{N}}$  uma seqüência de Cauchy. Como  $X$  é de dimensão finita,  $X$  é isomorfo a um  $\mathbb{K}^n$ , com  $\mathbb{K}$  real ou complexo, e  $n = \dim(X)$ . Seja

$J : X \rightarrow \mathbb{K}^n$  esse isomorfismo. Graças à estimativa (1.1) a ser feita na demonstração do Teorema 1.13, existem  $\mu, C > 0$  tais que

$$\mu \|J(\mathbf{u})\|_\infty \leq \|\mathbf{u}\|_X \leq C \|J(\mathbf{u})\|_\infty,$$

para todo  $\mathbf{u} \in X$ . Logo,

$$\|J(\mathbf{u}_k) - J(\mathbf{u}_l)\| = \|J(\mathbf{u}_k - \mathbf{u}_l)\| \leq \frac{1}{\mu} \|\mathbf{u}_k - \mathbf{u}_l\|_X \rightarrow 0, \quad k, l \rightarrow \infty.$$

Logo,  $\{J(\mathbf{u}_k)\}_k$  é uma sequência de Cauchy em  $\mathbb{K}^n$  com a norma do máximo. Como  $\mathbb{K}^n$  é completo, existe  $\mathbf{v} \in \mathbb{K}^n$  tal que

$$J(\mathbf{u}_k) \rightarrow \mathbf{v}.$$

Definindo  $\mathbf{u} = J^{-1}(\mathbf{v})$ , temos

$$\|\mathbf{u}_k - \mathbf{u}\|_X \leq C \|J(\mathbf{u}_k) - \mathbf{v}\|_\infty \rightarrow 0, \quad k \rightarrow \infty,$$

mostrando que  $\{\mathbf{u}_k\}_{k \in \mathbb{N}}$  é convergente e, portanto, que  $X$  é completo.  $\square$

Sendo  $\mathbf{u} \in A$  o limite de uma sequência  $\mathbf{u}_k$ , quando podemos garantir que  $\mathbf{u}_k$  está, eventualmente, contido em  $A$ ? E quando uma sequência  $\mathbf{u}_k$  em um conjunto  $A$  converge para um vetor  $\mathbf{u}$ , quando podemos garantir que o limite  $\mathbf{u}$  também está em  $A$ ? Para caracterizar essas propriedades, temos as noções de conjuntos *abertos* e *fechados*.

DEFINIÇÃO 1.5 (abertos e fechados). *Seja  $X$  um espaço vetorial normado real ou complexo, com norma  $\|\cdot\|$ . Um subconjunto  $A \subset X$  é dito **aberto** quando, para todo  $\mathbf{u} \in A$ , existe  $\delta > 0$  tal que todo elemento  $\mathbf{v} \in X$  com  $\|\mathbf{v} - \mathbf{u}\| < \delta$  está em  $A$ , i.e.*

$$\{\mathbf{v} \in X; \|\mathbf{v} - \mathbf{u}\| < \delta\} \subset A.$$

*Um conjunto  $A \subset X$  é dito **fechado** quando o seu complementar  $A^c = X \setminus A$  é aberto.*

Assim, temos o seguinte resultado.

TEOREMA 1.4. *Seja  $X$  um espaço vetorial normado real ou complexo e considere uma sequência  $(\mathbf{u}_k)_{k \in \mathbb{N}}$  em  $X$  convergindo para um vetor  $\mathbf{u} \in X$ .*

- (i) *Supondo  $\mathbf{u} \in O$  e  $O$  um conjunto aberto, então existe  $K \in \mathbb{N}$  tal que  $\mathbf{u}_k \in O$ , para todo  $k \geq K$ , ou seja, a sequência está eventualmente contida no conjunto aberto  $O$ .*
- (ii) *Supondo  $\mathbf{u}_k \in F$  e  $F$  um subconjunto fechado de  $X$ , então  $\mathbf{u} \in F$ , ou seja, o limite também está no conjunto fechado  $F$ .*

Os conjuntos abertos podem ser usados para caracterizar a convergência de uma sequência.

**TEOREMA 1.5.** *Seja  $X$  um espaço vetorial normado real ou complexo e considere uma sequência  $(\mathbf{u}_k)_{k \in \mathbb{N}}$  em  $X$  e um vetor  $\mathbf{u} \in X$ . Então  $\mathbf{u}_k$  converge para  $\mathbf{u}$  se, e somente se, para todo aberto  $O$  de  $X$  contendo  $\mathbf{u}$ , existe um inteiro  $K \in \mathbb{N}$  tal que  $\mathbf{u}_k \in O$ , para todo  $k \geq K$ , ou seja, a sequência eventualmente entra em qualquer aberto contendo o ponto limite.*

Por conta dessa caracterização, podemos definir *topologias* apenas em termos de conjuntos abertos e definir convergência, continuidade, etc., apenas em termos de abertos. O importante é destacar as propriedades fundamentais de conjuntos abertos apenas em termos de estrutura de conjuntos. Nesse sentido, observamos as seguintes propriedades dos abertos de um espaço vetorial normado.

**TEOREMA 1.6** (Topologia de um espaço vetorial normado). *Seja  $X$  um espaço vetorial normado real ou complexo. A **topologia** de  $X$  é o conjunto  $\mathcal{T}$  de todos os abertos de  $X$  e tem as seguintes propriedades.*

- (i) *O conjunto vazio é aberto (e fechado), i.e.  $\emptyset \in \mathcal{T}$ ;*
- (ii) *O espaço todo é aberto (e fechado), i.e.  $X \in \mathcal{T}$ ;*
- (iii) *A união arbitrária de abertos é aberta, i.e. se  $A_\lambda \in \mathcal{T}$ , para todo  $\lambda \in \Lambda$  em um conjunto  $\Lambda$  qualquer, então  $\bigcup_{\lambda \in \Lambda} A_\lambda \in \mathcal{T}$ ;*
- (iv) *A interseção finita de abertos é aberta, i.e. se  $A_1, \dots, A_n \in \mathcal{T}$ ,  $n \in \mathbb{N}$ , então  $\bigcap_{j=1, \dots, n} A_j \in \mathcal{T}$ .*

**DEMONSTRAÇÃO.** Deixamos como exercício para o leitor. □

Com base nessa caracterização, definimos *espaços topológicos* como os conjuntos  $X$  munidos de uma coleção  $\mathcal{T}$  de subconjuntos de  $X$  com as propriedades (i)-(iv) do **Teorema 1.6**. Mas não precisamos considerar essa generalização aqui, já que estamos tratando de espaços vetoriais de dimensão finita. Em dimensão infinita, esse conceito é bastante útil. Por exemplo, o espaço das distribuições (veja **Exemplo 4.1**) não é um espaço vetorial normado, mas é um *espaço vetorial topológico*, quando munido de uma topologia apropriada.

No caso de  $\mathbb{R}^n$ , todas as normas  $p$  geram a mesma topologia. Isso porque as normas são equivalentes. Vamos, primeiro, a esse conceito de equivalência de normas.

**DEFINIÇÃO 1.6** (Equivalência de normas). *Seja  $X$  um espaço vetorial real ou complexo e sejam  $\|\cdot\|_a$  e  $\|\cdot\|_b$  duas normas em  $X$ . As duas normas são ditas equivalentes quando existem constantes positivas  $c_1$  e  $c_2$  tais que*

$$\frac{1}{c_1} \|\mathbf{u}\|_a \leq \|\mathbf{u}\|_b \leq c_2 \|\mathbf{u}\|_a,$$

para todo  $\mathbf{u} \in X$ .

Normas equivalentes geram a mesma topologia.

TEOREMA 1.7. *Seja  $X$  um espaço vetorial e sejam  $\|\cdot\|_a$  e  $\|\cdot\|_b$  duas normas em  $X$ . Então um conjunto  $O$  é aberto em relação à norma  $\|\cdot\|_a$  se, e somente se, é aberto em relação à norma  $\|\cdot\|_b$ .*

No caso de  $\mathbb{R}^n$  e  $\mathbb{C}^n$ , todas as normas  $p$  são equivalentes.

TEOREMA 1.8. *Seja  $X = \mathbb{K}^n$ ,  $n \in \mathbb{N}$ , com  $\mathbb{K}$  real ou complexo. Então todas as normas  $p$ ,  $1 \leq p \leq \infty$ , são equivalentes entre si e consideramos a topologia de  $\mathbb{K}^n$  como sendo a topologia gerada por essas normas.*

DEMONSTRAÇÃO. Observe que a equivalência entre normas de um mesmo espaço vetorial é uma relação de equivalência. Assim, para compararmos duas normas  $p_1$  e  $p_2$  quaisquer, podemos comparar cada uma delas com uma mesma norma  $q$ . Vamos usar  $q = \infty$  para isso e comparar uma norma  $p$  qualquer em  $1 \leq p < \infty$  com  $q = \infty$ .

Seja  $\mathbf{u} = (x_1, \dots, x_n) \in \mathbb{K}^n$ . Temos

$$|x_j|^p \leq |x_1|^p + \dots + |x_n|^p,$$

para todo  $j = 1, \dots, n$ , de modo que

$$\begin{aligned} \|\mathbf{u}\|_\infty &= \max_{j=1, \dots, n} |x_j| = \left( \max_{j=1, \dots, n} |x_j|^p \right)^{1/p} \\ &\leq (|x_1|^p + \dots + |x_n|^p)^{1/p} = \|\mathbf{u}\|_p. \end{aligned}$$

Por outro lado, temos

$$\begin{aligned} \|\mathbf{u}\|_p &= (|x_1|^p + \dots + |x_n|^p)^{1/p} \leq \left( n \max_{j=1, \dots, n} \{|x_j|^p\} \right)^{1/p} \\ &\leq n^{1/p} \max_{j=1, \dots, n} |x_j| = n^{1/p} \|\mathbf{u}\|_\infty. \end{aligned}$$

Portanto,

$$\frac{1}{n^{1/p}} \|\mathbf{u}\|_p \leq \|\mathbf{u}\|_\infty \leq \|\mathbf{u}\|_p,$$

para todo  $\mathbf{u} \in X$ , mostrando que a norma  $p$ , com  $1 \leq p < \infty$ , é equivalente à norma  $q = \infty$ .

Como dito acima, dadas normas  $p_1, p_2$ ,  $1 \leq p_1, p_2 < \infty$ , temos, usando a equivalência acima, que

$$\frac{1}{n^{1/p_1}} \|\mathbf{u}\|_{p_1} \leq \|\mathbf{u}\|_\infty \leq \|\mathbf{u}\|_{p_2} \leq n^{1/p_2} \|\mathbf{u}\|_\infty \leq n^{1/p_2} \|\mathbf{u}\|_{p_1},$$

completando a equivalência entre normas  $p_1$  e  $p_2$  quaisquer.  $\square$

OBSERVAÇÃO 1.5. Podemos definir outras métricas e outras topologias em  $\mathbb{R}^n$ . Por exemplo, em um espaço vetorial qualquer, a função  $d(\mathbf{u}, \mathbf{v}) = 1$ , se  $\mathbf{u} \neq \mathbf{v}$  e  $d(\mathbf{u}, \mathbf{v}) = 0$ , para  $\mathbf{u} = \mathbf{v}$ , define uma métrica que nos dá o que chamamos de *topologia discreta*, em que todos os subconjuntos de  $X$  são abertos e fechados. Essa topologia, no entanto, não é proveniente de uma norma.

OBSERVAÇÃO 1.6. Métricas diferentes podem gerar topologias diferentes, mesmo em dimensão finita, como mostra o [Observação 1.5](#). Mas, no caso de normas, vimos, no [Teorema 1.8](#), que todas as normas  $p$  em  $\mathbb{R}^n$  e  $\mathbb{C}^n$  são equivalentes e geram a mesma topologia. Esse resultado é mais geral ainda, como veremos adiante. Em qualquer espaço de dimensão finita, todas as normas são equivalentes entre si. Para provarmos esse resultado mais geral, precisamos da noção de compacidade.

Com o conceito de norma, podemos definir conjuntos especiais que são equidistantes a um determinado ponto.

DEFINIÇÃO 1.7. *Seja  $X$  um espaço vetorial normado real ou complexo, com norma  $\|\cdot\|$ . Uma **bola fechada** de raio  $r \geq 0$  e centro em  $\mathbf{u}_0$  é o conjunto*

$$B_r(\mathbf{u}_0) = \{\mathbf{u} \in X; \|\mathbf{u} - \mathbf{u}_0\| \leq r\},$$

enquanto que a **bola aberta** é definida por

$$B_r(\mathbf{u}_0)^\circ = \{\mathbf{u} \in X; \|\mathbf{u} - \mathbf{u}_0\| < r\}.$$

A **esfera** de raio  $r$  e centrada em  $\mathbf{u}_0$  é dada por

$$S_r(\mathbf{u}_0) = \{\mathbf{u} \in X; \|\mathbf{u} - \mathbf{u}_0\| = r\}.$$

Um resultado fundamental de análise que usaremos aqui e que anunciamos sem demonstração é o da compacidade de subconjuntos fechados limitados em  $\mathbb{R}^n$ . Há dois resultados importantes nesse sentido, ligados a conceitos diferentes de compacidade. O de *Bolzano-Weierstrass*, que garante que toda sequência limitada em  $\mathbb{R}^n$  possui uma subsequência convergente, e que implica em que todo subconjunto fechado e limitado de  $\mathbb{R}^n$  é *sequencialmente compacto* (ou seja, toda sequência em um conjunto fechado e limitado  $F$  de  $\mathbb{R}^n$  possui uma subsequência que converge para algum elemento de  $F$ ). E o de *Heine-Borel*, que garante que todo subconjunto fechado e limitado  $F$  de  $\mathbb{R}^n$  é *compacto*, no sentido de que toda cobertura  $A \subset \cup_{\lambda \in \Lambda} O_\lambda$  de  $A$  por conjuntos abertos  $O_\lambda$  de uma família arbitrária  $\Lambda$  possui uma subcobertura finita  $A \subset \cup_{j=1, \dots, n} O_{\lambda_j}$ ,  $\lambda_j \in \Lambda$ ,  $n \in \mathbb{N}$ . Este último conceito é a generalização de compacidade sequencial para espaços topológicos mais gerais. No caso de espaço métricos e, em particular, de espaços vetoriais normados, os dois conceitos coincidem e não precisamos nos preocupar com essa distinção na definição de compacidade. E o mesmo vale para

$\mathbb{C}^n$  ou qualquer espaço vetorial complexo de dimensão finita. Basta considerarmos as suas partes real e imaginária como  $\mathbb{R}^n$  e usar os resultados mencionados acima.

DEFINIÇÃO 1.8 (Compacidade em espaços vetoriais normados). *Um subconjunto  $A$  de um espaço vetorial normado  $X$  real ou complexo é dito **compacto** quando toda sequência  $(\mathbf{u}_n)_{n \in \mathbb{N}} \subset A$  de elementos em  $A$  possui uma subsequência convergente para um elemento de  $A$ .*

Agora, para os teoremas de compacidade, observamos que a norma nos permite classificar subconjuntos entre limitados e ilimitados.

DEFINIÇÃO 1.9. *Seja  $X$  um espaço vetorial normado, com norma  $\|\cdot\|$ . Um subconjunto  $A$  de  $X$  é dito **limitado** quando existe uma constante  $C > 0$  tal que*

$$\|\mathbf{u}\| \leq C, \quad \forall \mathbf{u} \in A.$$

Um subconjunto  $A$  é limitado se, e somente se, ele está contido em uma bola centrada na origem, por exemplo.

De posse do conceito de conjunto limitado, podemos enunciar o Teorema de Bolzano-Weierstrass.

TEOREMA 1.9 (de Bolzano-Weierstrass). *Considere  $\mathbb{R}^n$  munido da topologia associada a uma norma  $p$  qualquer,  $1 \leq p \leq \infty$ . Então toda subsequência limitada  $(\mathbf{u}_k)_{k \in \mathbb{N}}$  possui uma subsequência convergente.*

Juntando com os conceitos de conjunto fechado e de conjunto compacto, temos o Teorema de Heine-Borel.

TEOREMA 1.10 (de Heine-Borel). *Considere  $\mathbb{R}^n$  munido da topologia associada a uma norma  $p$  qualquer,  $1 \leq p \leq \infty$ . Então todo subconjunto fechado e limitado de  $\mathbb{R}^n$  é compacto.*

Identificando  $\mathbb{C}^n$  com  $\mathbb{R}^n \times \mathbb{R}^n$  via partes real e imaginária de um vetor complexo, os resultados acima também se aplicam ao  $\mathbb{C}^n$ .

TEOREMA 1.11 (Compacidade nos complexos). *Considere  $\mathbb{C}^n$  munido da topologia associada a uma norma  $p$  qualquer,  $1 \leq p \leq \infty$ . Então todo subconjunto fechado e limitado de  $\mathbb{C}^n$  é compacto.*

Um outro resultado fundamental em análise e que nos será útil a seguir é o de que funções contínuas definidas em conjuntos compactos são limitadas e assumem valores máximos e mínimos. Também anunciamos esse resultado sem demonstração.

TEOREMA 1.12. *Seja  $A \subset \mathbb{K}^n$  um subconjunto compacto de  $\mathbb{K}^n$  em um corpo  $\mathbb{K}$  real ou complexo e seja  $f : A \rightarrow \mathbb{R}$  uma função contínua. Então  $f$  é limitada e possui pontos de máximo e de mínimo, i.e. existem  $\mathbf{u}_{\min}$  e  $\mathbf{u}_{\max}$  em  $A$  tais que*

$$f(\mathbf{u}_{\min}) = \min_{\mathbf{u} \in A} f(\mathbf{u}), \quad f(\mathbf{u}_{\max}) = \max_{\mathbf{u} \in A} f(\mathbf{u}).$$

Essa compacidade pode ser usada para mostrar que não apenas as normas  $p$  mas, também, quaisquer normas em  $\mathbb{R}^n$  e  $\mathbb{C}^n$  são equivalentes. Mais ainda, quaisquer normas em espaços vetoriais de dimensão finita são equivalentes entre si.

TEOREMA 1.13. *Seja  $X$  um espaço vetorial real ou complexo de dimensão finita. Então quaisquer duas normas  $\|\cdot\|_a$  e  $\|\cdot\|_b$  em  $X$  são equivalentes entre si.*

DEMONSTRAÇÃO. Se  $X$  é trivial, então  $\mathbf{0}$  é o seu único elemento e  $\|\mathbf{0}\|_a = \|\mathbf{0}\|_b = 0$ , de modo que as normas são trivialmente equivalentes. Suponha  $X$  de dimensão  $n \in \mathbb{N}$ . Seja  $\mathcal{b} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  uma base de  $X$  e  $\mathcal{b}' = \{\mathbf{f}_1, \dots, \mathbf{f}_n\}$  a base dual. Considere o isomorfismo

$$J : X \rightarrow \mathbb{K}^n$$

dado por

$$J(\mathbf{u}) = (\langle \mathbf{f}_1, \mathbf{u} \rangle, \dots, \langle \mathbf{f}_n, \mathbf{u} \rangle) \in \mathbb{K}^n,$$

onde  $\mathbb{K}$  é real ou complexo, dependendo do caso. A sua inversa é dada por

$$J^{-1}((x_1, \dots, x_n)) = x_1 \mathbf{w}_1 + \dots + x_n \mathbf{w}_n.$$

Usando a desigualdade triangular e a homogeneidade da norma, temos

$$\begin{aligned} \|J^{-1}((x_1, \dots, x_n))\|_a &= \|x_1 \mathbf{w}_1 + \dots + x_n \mathbf{w}_n\|_a \leq |x_1| \|\mathbf{w}_1\|_a + \dots + |x_n| \|\mathbf{w}_n\|_a \\ &\leq \left( \sum_j \|\mathbf{w}_j\|_a \right) \max\{|x_1|, \dots, |x_n|\}, \end{aligned}$$

ou seja,

$$\|J^{-1}\mathbf{v}\|_a \leq C_a \|\mathbf{v}\|_\infty,$$

para todo  $\mathbf{v} = (x_1, \dots, x_n) \in \mathbb{K}^n$ , onde

$$C_a = \sum_j \|\mathbf{w}_j\|_a.$$

Em particular, isso mostra a continuidade de  $J^{-1}$  como função de  $\mathbb{K}^n$  em  $X$ .

Considerando a esfera unitária

$$S = \{\mathbf{v} \in \mathbb{K}^n; \|\mathbf{v}\|_\infty = 1\},$$

temos que  $S$  é um subconjunto fechado e limitado em  $\mathbb{K}^n$ , portanto compacto, graças ao [Teorema 1.10](#) de Heine-Borel (ou [Teorema 1.11](#) no caso complexo).

Agora, considere a função  $f : \mathbb{K}^n \rightarrow \mathbb{R}$  dada por  $f(\mathbf{v}) = \|\mathbf{J}^{-1}\mathbf{v}\|_a$ . Temos que  $f$  é uma função contínua, visto que, se  $\mathbf{v}_k \rightarrow \mathbf{v}$ , então, pela linearidade de  $\mathbf{J}^{-1}$  e pelo continuidade da norma garantida pelo [Teorema 1.1](#),

$$|f(\mathbf{v}_k) - f(\mathbf{v})| \leq \|\mathbf{J}^{-1}(\mathbf{v}_k - \mathbf{v})\|_a \leq C\|\mathbf{v}_k - \mathbf{v}\| \rightarrow 0.$$

Considerando  $f$  restrito ao compacto  $S$ , deduzimos, do [Teorema 1.12](#), que  $f$  possui pelo menos um ponto de mínimo  $\mathbf{v}_0 \in S$ ,

$$\mu_a = f(\mathbf{v}_0) = \min_{\mathbf{v} \in S} f(\mathbf{v}).$$

Temos  $\mu_a = f(\mathbf{v}_0) = \|\mathbf{J}^{-1}(\mathbf{v}_0)\|_a \geq 0$ . Caso  $\mu_a = 0$ , então, pela positividade da norma,  $\mathbf{J}^{-1}(\mathbf{v}_0) = \mathbf{0}$ . Como  $\mathbf{J}$  é um isomorfismo, então  $\mathbf{v}_0 = \mathbf{0}$  também. Mas  $\mathbf{v}_0 \in S$ , o que significa que  $\|\mathbf{v}_0\|_\infty = 1$ . Logo,  $\mathbf{v}_0$  não pode ser  $\mathbf{0}$  e  $\mu_a$  não pode ser nulo. Portanto,  $\mu_a > 0$ . Ou seja,

$$\|\mathbf{J}^{-1}\mathbf{v}\|_a \geq \mu_a > 0,$$

para todo  $\mathbf{v} \in S$ . Seja, agora,  $\mathbf{u} \in X$  não nulo arbitrário. Então

$$\mathbf{v} = \frac{1}{\|\mathbf{J}(\mathbf{u})\|_\infty} \mathbf{J}(\mathbf{u}) \in S.$$

Logo,

$$\|\mathbf{J}^{-1}\mathbf{v}\|_a \geq \mu_a.$$

Usando a linearidade de  $\mathbf{J}^{-1}$  e a homogeneidade da norma, temos

$$\mu_a \leq \frac{1}{\|\mathbf{J}(\mathbf{u})\|_\infty} \|\mathbf{J}^{-1}\mathbf{J}(\mathbf{u})\|_a = \frac{1}{\|\mathbf{J}(\mathbf{u})\|_\infty} \|\mathbf{u}\|_a.$$

Ou seja,

$$\mu_a \|\mathbf{J}(\mathbf{u})\|_\infty \leq \|\mathbf{u}\|_a,$$

para  $\mathbf{u} \in X$ ,  $\mathbf{u} \neq \mathbf{0}$ . Juntando com a estimativa anterior para  $\mathbf{v} = \mathbf{J}(\mathbf{u})$  e incluindo o caso  $\mathbf{u} = \mathbf{0}$ , já que agora não estamos mais dividindo pela norma, chegamos a

$$\mu_a \|\mathbf{J}(\mathbf{u})\|_\infty \leq \|\mathbf{u}\|_a \leq C_a \|\mathbf{J}(\mathbf{u})\|_\infty, \quad (1.1)$$

para todo  $\mathbf{u} \in X$ .

Repetindo o argumento para a norma  $\|\cdot\|_b$ , obtemos

$$\mu_b \|\mathbf{J}(\mathbf{u})\|_\infty \leq \|\mathbf{u}\|_b \leq C_b \|\mathbf{J}(\mathbf{u})\|_\infty,$$

para  $\mu_b, C_b > 0$  apropriados. Com isso, chegamos à equivalência

$$\frac{\mu_a}{C_b} \|\mathbf{u}\|_b \leq \mu_a \|\mathbf{J}(\mathbf{u})\|_\infty \leq \|\mathbf{u}\|_a \leq C_a \|\mathbf{J}(\mathbf{u})\|_\infty \leq \frac{C_a}{\mu_b} \|\mathbf{u}\|_b,$$

para todo  $\mathbf{u} \in X$ . □

Graças a esse resultado, as propriedades puramente topológicas de um espaço vetorial normado independem da escolha da norma. Em dimensão infinita, no entanto, as propriedades topológicas dependem de maneira não trivial da norma escolhida. Além disso, temos topologias bastante importantes que nem são provenientes de norma. Várias coisas podem funcionar em uma topologia e não funcionar em outra (continuidade de operadores, compacidade de certos subconjuntos, existência e/ou unicidade de soluções de determinadas equações, etc.). Mas isso é outro assunto. Aqui, vamos nos aprofundar em espaços de dimensão finita, onde essas diferenças não acontecem.

Os resultados mencionados nessa seção são essencialmente para referência. Eles são explorados mais profundamente em cursos de Análise Real, Topologia e Análise Funcional.

### 1.3. Norma de operadores e continuidade de transformações lineares.

Com a noção de distância em espaços vetoriais, podemos analisar a continuidade ou não de transformação lineares. Nesse ponto, a dimensão do espaço tem um papel fundamental. Em dimensão finita, veremos que todas as transformações lineares são contínuas. Mas em dimensão infinita, podemos ter transformações lineares que não são contínuas, como mostra o seguinte exemplo.

EXEMPLO 1.4. Seja

$$X = c_{00}(\mathbb{R}) = \{\mathbf{u} = (x_j)_{j \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}; \exists n = n(\mathbf{u}) \in \mathbb{N}, x_j = 0, \forall j \geq n\},$$

o espaço das sequências que eventualmente se anulam, munido da norma do máximo

$$\|\mathbf{u}\|_X = \|\mathbf{u}\|_{\infty} = \sup_{j \in \mathbb{N}} |x_j|,$$

e seja

$$Y = \ell^1(\mathbb{R}) = \{\mathbf{u} = (x_j)_{j \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}; \sum_{n \in \mathbb{N}} |x_j| < \infty\}$$

o espaço das sequências absolutamente convergentes, munido da norma da soma

$$\|\mathbf{u}\|_Y = \|\mathbf{u}\|_1 = \sum_{j \in \mathbb{N}} |x_j|.$$

Ambos são espaços vetoriais normados. Agora considere a transformação linear

$$T(\mathbf{u}) = (jx_j)_{j \in \mathbb{N}}, \quad \mathbf{u} = (x_j)_{j \in \mathbb{N}} \in c_{00}(\mathbb{R}).$$

Para cada  $\mathbf{u} \in c_{00}(\mathbb{R})$ , como a sequência  $\mathbf{u} = (x_j)_{j \in \mathbb{N}}$  eventualmente se anula, i.e. existe  $n = n(\mathbf{u})$  tal que  $x_j = 0$ , para todo  $j \geq n$ , então

$$\|\mathbb{T}(\mathbf{u})\|_1 = \sum_{j \in \mathbb{N}} |jx_j| = \sum_{j=1}^{n(\mathbf{u})} j|x_j| < \infty,$$

de modo que  $\mathbb{T}$  está bem definido e é, claramente, uma transformação linear. Portanto,  $\mathbb{T} \in \mathcal{L}(X, Y)$ . No entanto,  $\mathbb{T}$  não é contínua em relação a essas normas. De fato, considere a sequência

$$\mathbf{u}_n = \begin{cases} 1/n, & 1 \leq j \leq n, \\ 0, & j > n. \end{cases}$$

Cada  $\mathbf{u}_n \in X$ , com

$$\|\mathbf{u}_n\|_X = \|\mathbf{u}_n\|_\infty = \frac{1}{n} \rightarrow 0,$$

ou seja,  $\mathbf{u}_n \rightarrow \mathbf{0}$  em  $X$ . Por outro lado,

$$\|\mathbb{T}(\mathbf{u}_n)\|_Y = \sum_{j=1}^n \frac{j}{n} = \frac{1}{n} \sum_{j=1}^n j = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2} \rightarrow \infty.$$

Ou seja,  $\mathbb{T}(\mathbf{u}_n)$  não converge para  $\mathbb{T}(\mathbf{0}) = \mathbf{0}$ , de modo que  $\mathbb{T}$  não é contínua.

Uma característica de transformações lineares é que, pela linearidade da transformação, em combinação com a estrutura linear do espaço e a homogeneidade da norma, a continuidade de uma transformação linear pode ser analisada apenas pela continuidade na origem. Mais ainda, essa continuidade na origem segue de maneira ainda mais simples, bastando que a transformação seja limitada em conjuntos limitados. Nesse sentido, temos o seguinte resultado.

**TEOREMA 1.14.** *Sejam  $X$  e  $Y$  espaços vetoriais sobre um mesmo corpo, real ou complexo, e munidos de normas  $\|\cdot\|_X$  e  $\|\cdot\|_Y$ , respectivamente. Dada uma transformação linear  $\mathbb{T} \in \mathcal{L}(X, Y)$ , temos as seguintes equivalências*

- (i)  $\mathbb{T}$  é contínua em todos os pontos de  $X$ ;
- (ii)  $\mathbb{T}$  é contínua na origem;
- (iii)  $\mathbb{T}$  é limitada em  $B_r = \{\mathbf{u} \in X; \|\mathbf{u}\|_X \leq r\}$ , para algum  $r > 0$ .
- (iv)  $\mathbb{T}$  é limitada em qualquer conjunto limitado  $B_r$ ,  $r > 0$ ;

**DEMONSTRAÇÃO.** É imediato observar que  $\mathbb{T}$  ser contínua em  $X$  significa ser contínua em cada ponto de  $X$  e, em particular ser contínua na origem. Ou seja, (i)

$\Rightarrow$  (ii). Suponha, agora,  $\mathbf{T}$  contínua na origem. Como  $\mathbf{T}\mathbf{0} = \mathbf{0}$ , a continuidade na origem diz que, dado  $\varepsilon > 0$ , existe  $\delta > 0$  tal que

$$\|\mathbf{T}\mathbf{v}\|_Y < \varepsilon,$$

para todo  $\mathbf{v} \in X$  com

$$\|\mathbf{v}\|_X < \delta.$$

Seja, então,  $\mathbf{u} \in X$  e considere  $\tilde{\mathbf{u}} \in X$  arbitrário com

$$\|\tilde{\mathbf{u}} - \mathbf{u}\|_X < \delta.$$

Então, aplicando o fato anterior a  $\mathbf{v} = \tilde{\mathbf{u}} - \mathbf{u}$ , obtemos, visto que  $\|\mathbf{v}\|_X < \delta$ , que

$$\|\mathbf{T}\tilde{\mathbf{u}} - \mathbf{T}\mathbf{u}\|_Y = \|\mathbf{T}(\mathbf{v})\| < \varepsilon,$$

mostrando a continuidade de  $\mathbf{T}$  em  $\mathbf{u}$ . Isso nos dá a equivalência (i)  $\Leftrightarrow$  (ii)

Suponha, agora,  $\mathbf{T}$  contínua na origem. Então, dado  $\varepsilon > 0$ , existe  $\delta > 0$  tal que

$$\|\mathbf{T}\mathbf{v}\|_Y < \varepsilon,$$

para todo  $\mathbf{v} \in X$  com

$$\|\mathbf{v}\|_X < \delta.$$

Isso significa que

$$\sup_{\mathbf{v} \in B_r} \|\mathbf{T}\mathbf{v}\|_Y < \varepsilon,$$

para  $r < \delta$ , ou seja,  $\mathbf{T}$  é limitado em algum  $B_{r_0}$ , provando (iii). Além disso, usando essa limitação, temos, para qualquer  $r > 0$ , que se  $\mathbf{v} \in B_r$ , então  $\mathbf{w} = (r_0/r)\mathbf{v} \in B_{r_0}$ , de modo que

$$\begin{aligned} \sup_{\mathbf{v} \in B_r} \|\mathbf{T}\mathbf{v}\|_Y &= \frac{r}{r_0} \sup_{\mathbf{v} \in B_r} \frac{r_0}{r} \|\mathbf{T}\mathbf{v}\|_Y = \frac{r}{r_0} \sup_{\mathbf{v} \in B_r} \left\| \frac{r_0}{r} \mathbf{T}\mathbf{v} \right\|_Y \\ &= \frac{r}{r_0} \sup_{\mathbf{w} \in B_{r_0}} \|\mathbf{T}\mathbf{w}\|_Y \leq \frac{r\varepsilon}{r_0} < \infty, \end{aligned}$$

mostrando que  $\mathbf{T}$  é limitado em  $B_r$ , para  $r > 0$  arbitrário. Isso prova (iv). Obviamente, (iv) implica em (iii). Ou seja, já temos (i)  $\Leftrightarrow$  (ii)  $\Rightarrow$  (iii)  $\Leftrightarrow$  (iv).

Resta mostrar que (iii) implica em (ii). Para isso, vamos assumir que  $\mathbf{T}$  é limitada em  $B_{r_0}$  para algum  $r_0$ , digamos

$$\ell_0 = \sup\{\|\mathbf{T}\mathbf{v}\|_Y; \mathbf{v} \in B_{r_0}\} < \infty.$$

Dado  $\varepsilon > 0$ , basta, então, tomar

$$\delta = \frac{\varepsilon r_0}{\ell_0},$$

de modo que, para

$$\|\mathbf{v}\|_X < \delta,$$

temos

$$\left\| \frac{r_0}{\delta} \mathbf{v} \right\| = \frac{r_0}{\delta} \|\mathbf{v}\| < r_0,$$

ou seja,

$$\frac{r_0}{\delta} \mathbf{v} \in B_{r_0},$$

e, com isso,

$$\|\mathbf{T}\mathbf{v}\|_Y = \frac{\delta}{r_0} \|\mathbf{T} \left( \frac{r_0}{\delta} \mathbf{v} \right)\| \leq \frac{\delta}{r_0} \ell_0 = \varepsilon,$$

provando a continuidade na origem, i.e. provando que (iii)  $\Rightarrow$  (ii). Isso completa a demonstração.  $\square$

Inspirado pelo resultado do Teorema 1.14, definimos o conceito de transformação limitada.

DEFINIÇÃO 1.10. *Sejam  $X$  e  $Y$  espaços vetoriais normados reais ou complexos, munidos de normas  $\|\cdot\|_X$  e  $\|\cdot\|_Y$ , respectivamente. Seja  $\mathbf{T} \in \mathcal{L}(X, Y)$  uma transformação linear de  $X$  em  $Y$ . Dizemos que  $\mathbf{T}$  é **limitada** quando  $\|\mathbf{T}\mathbf{u}\|_Y$  é limitado uniformemente em qualquer subconjunto limitado de  $X$ , i.e. para todo  $r \geq 0$ , existe  $C \geq 0$  tal que*

$$\|\mathbf{T}\mathbf{u}\|_Y \leq C, \quad \forall \mathbf{u}, \|\mathbf{u}\|_X \leq r.$$

Com isso, é fácil provar a continuidade de transformações lineares associadas a matrizes.

TEOREMA 1.15 (Continuidade de transformações associadas a matrizes). *Seja  $\mathbf{A} = (a_{ij})_{i,j=1}^{m,n} \in \mathbb{K}^{m \times n}$  uma matriz em um corpo  $\mathbb{K}$  real ou complexo. Então*

$$\|\mathbf{A}\mathbf{u}\|_1 \leq \left( \sum_{i=1}^m \max_{j=1, \dots, n} |a_{ij}| \right) \|\mathbf{u}\|_1, \quad (1.2)$$

para todo  $\mathbf{u} \in \mathbb{R}^n$ , onde  $\|\cdot\|_1$  é a norma  $\ell^1$ , da soma, em  $\mathbb{R}^n$  e em  $\mathbb{R}^m$ . Portanto, a transformação  $\mathbf{T}_{\mathbf{A}}$  associada a  $\mathbf{A}$  é limitada e contínua de  $\mathbb{R}^n$  em  $\mathbb{R}^m$ , sob quaisquer normas, visto que as normas em dimensão finita são equivalentes.

DEMONSTRAÇÃO. Usando que  $\mathbf{A}\mathbf{u} = \left( \sum_{j=1}^n a_{ij} x_j \right)_{i=1}^m$ , temos

$$\|\mathbf{A}\mathbf{u}\|_1 = \sum_{i=1}^m \left| \sum_{j=1}^n a_{ij} x_j \right| \leq \sum_{i=1}^m \sum_{j=1}^n \max_{k=1, \dots, n} |a_{ik}| |x_j| \leq \left( \sum_{i=1}^m \max_{j=1, \dots, n} |a_{ij}| \right) \|\mathbf{u}\|_1.$$

□

Usando o resultado acima, a equivalência das normas em dimensão finita e o fato de que toda transformação linear entre espaços de dimensão finita pode ser representada por uma matriz, obtemos a continuidade de transformações lineares entre espaços de dimensão finita quaisquer.

**TEOREMA 1.16** (Continuidade de transformações lineares em dimensão finita). *Sejam  $X$  e  $Y$  espaços vetoriais normados reais ou complexos de dimensão finita. Então toda transformação linear  $T \in \mathcal{L}(X, Y)$  de  $X$  em  $Y$  é contínua.*

**DEMONSTRAÇÃO.** Graças a **Teorema 1.14**, basta mostrar que  $T$  é limitada. Escolhendo bases  $\alpha$  e  $\beta$  de  $X$  e  $Y$ , respectivamente, temos que

$$[T\mathbf{u}]_{\beta} = [T]_{\beta}^{\alpha} [\mathbf{u}]_{\alpha},$$

onde  $[T]_{\beta}^{\alpha}$  é uma matriz. Graças ao **Teorema 1.15**, existe uma constante  $C \geq 0$  tal que

$$\|[T\mathbf{u}]_{\beta}\|_1 = \|[T]_{\beta}^{\alpha} [\mathbf{u}]_{\alpha}\|_1 \leq C \|[u]_{\alpha}\|_1.$$

Conforme feito na demonstração do **Teorema 1.13**, mais precisamente em (1.1), existem constantes  $C_Y \geq 0$  e  $\mu_X > 0$  tais que

$$\|T\mathbf{u}\|_Y \leq C_Y \|[T\mathbf{u}]_{\beta}\|_{\infty},$$

e

$$\|[u]_{\alpha}\|_{\infty} \leq \frac{1}{\mu_X} \|\mathbf{u}\|_X.$$

Usando a equivalência entre as normas  $p$  garantida pelo **Teorema 1.8**, com a constante explícita obtida na demonstração, temos

$$\|[T\mathbf{u}]_{\beta}\|_{\infty} \leq \|[T\mathbf{u}]_{\beta}\|_1$$

e

$$\|[u]_{\alpha}\|_1 \leq n \|[u]_{\alpha}\|_{\infty},$$

onde  $n$  é a dimensão de  $Y$ . Juntando as desigualdades, obtemos

$$\begin{aligned} \|T\mathbf{u}\|_Y &\leq C_Y \|[T\mathbf{u}]_{\beta}\|_{\infty} \leq C_Y \|[T\mathbf{u}]_{\beta}\|_1 \\ &\leq C_Y C \|[u]_{\alpha}\|_1 \\ &\leq n C_Y C \|[u]_{\alpha}\|_{\infty} \leq \frac{n C_Y C}{\mu_X} \|\mathbf{u}\|_X. \end{aligned}$$

Como isso vale para  $\mathbf{u} \in X$  arbitrário, obtemos que  $T$  é limitada e, portanto, contínua. □

No **Teorema 1.15**, vimos que, para toda matriz em  $\mathbb{K}^{m \times n}$ , com  $\mathbb{K}$  real ou complexo, vale a desigualdade (1.2). O somatório do máximo dos coeficientes da matriz, que aparece nessa desigualdade, define, de fato, uma norma no próprio espaço  $\mathbb{K}^{m \times n}$  das matrizes. Mais geralmente, para  $1 \leq p, q < \infty$ , podemos definir

$$\|\mathbf{A}\|_{p,q} = \left( \sum_{j=1}^n \left( \sum_{i=1}^m |a_{ij}|^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}}, \quad (1.3)$$

que são também normas em  $\mathbb{K}^{m \times n}$ . Deixamos isso como exercício. Versões para  $p$  e/ou  $q$  infinitos também podem ser definidas.

Uma propriedade dessas normas de matrizes é a seguinte.

**TEOREMA 1.17.** *Seja  $\mathbb{K}$  o corpo dos reais ou dos complexos e sejam  $p, q$  reais tais que  $1 \leq p, q \leq \infty$ . Então*

$$\|\mathbf{A}\mathbf{u}\|_q \leq \|\mathbf{A}\|_{p',q} \|\mathbf{u}\|_p, \quad (1.4)$$

para todo  $\mathbf{u} \in \mathbb{K}^n$ , onde  $1 \leq p' \leq \infty$  é tal que  $1/p + 1/p' = 1$ , com  $p' = 1$ , se  $p = \infty$  e  $p' = \infty$ , se  $p = 1$ .

**DEMONSTRAÇÃO.** Escreva  $\mathbf{A} = (a_{ij})_{i,j=1}^{i=m,j=n}$ . Temos  $(\mathbf{A}\mathbf{u}) = \left( \sum_{j=1}^n a_{ij}x_j \right)$ , para  $\mathbf{u} = (x_1, \dots, x_n)$ . Assim,

$$\|\mathbf{A}\mathbf{u}\|_q^q = \sum_{i=1}^m \left| \sum_{j=1}^n a_{ij}x_j \right|^q.$$

A desigualdade de Hölder nos diz que

$$\left| \sum_{j=1}^n a_{ij}x_j \right| \leq \left( \sum_{j=1}^n |a_{ij}|^{p'} \right)^{1/p'} \left( \sum_{j=1}^n |x_j|^p \right)^{1/p}.$$

Assim,

$$\|\mathbf{A}\mathbf{u}\|_q^q \leq \sum_{i=1}^m \left( \sum_{j=1}^n |a_{ij}|^{p'} \right)^{q/p'} \left( \sum_{j=1}^n |x_j|^p \right)^{q/p}.$$

Como o somatório de  $|x_j|^p$  independe de  $i$ , temos

$$\|\mathbf{A}\mathbf{u}\|_q^q \leq \left( \sum_{i=1}^m \left( \sum_{j=1}^n |a_{ij}|^{p'} \right)^{q/p'} \right) \left( \sum_{j=1}^n |x_j|^p \right)^{q/p}.$$

Extraindo a raiz  $q$ , obtemos

$$\|\mathbf{A}\mathbf{u}\|_q \leq \left( \sum_{i=1}^m \left( \sum_{j=1}^n |a_{ij}|^{p'} \right)^{q/p'} \right)^{1/q} \|\mathbf{u}\|_p = \|\mathbf{A}\|_{p',q} \|\mathbf{u}\|_p,$$

provando a desigualdade.  $\square$

OBSERVAÇÃO 1.7. A norma (1.3) no caso  $p = q = 2$  nos dá uma norma especial que leva o nome de *norma de Frobenius*, denotada por  $\|\cdot\|_F$ , dada simplesmente por

$$\|\mathbf{A}\|_F^2 = \sum_{j=1}^n \sum_{i=1}^m a_{ij}^2.$$

Essa norma é bastante útil na prática, nos dando uma boa estimativa para a norma  $\ell^2$  (veja (1.4)):

$$\|\mathbf{A}\mathbf{u}\|_2 \leq \|\mathbf{A}\|_F \|\mathbf{u}\|_2,$$

para todo  $\mathbf{u} \in \mathbb{R}^n$ . Munindo  $\mathbb{R}^n$  com a norma euclidiana, podemos obter a equivalência

$$\|\mathbf{A}\|_{\mathcal{L}(\mathbb{R}^n)} \leq \|\mathbf{A}\|_F \leq \sqrt{n} \|\mathbf{A}\|_{\mathcal{L}(\mathbb{R}^n)},$$

onde

$$\|\mathbf{A}\|_{\mathcal{L}(\mathbb{R}^n)} = \sup_{\mathbf{u} \neq \mathbf{0}} \frac{\|\mathbf{A}\mathbf{u}\|_2}{\|\mathbf{u}\|_2}.$$

A estimativa (1.4) nos faz pensar se é possível obter uma constante mais justa, ou ótima, limitando  $\|\mathbf{A}\mathbf{u}\|_q$  em termos de  $\|\mathbf{u}\|_p$ . E se algo assim pode ser obtido entre normas quaisquer. De fato, podemos definir essa constante ótima via

$$C = \max_{\mathbf{u} \neq \mathbf{0}} \frac{\|\mathbf{A}\mathbf{u}\|_q}{\|\mathbf{u}\|_p}.$$

Acontece que essa definição nos dá, também, uma norma. De fato, temos o seguinte resultado.

TEOREMA 1.18. *Sejam  $X$  e  $Y$  espaços vetoriais normados sobre um mesmo corpo  $\mathbb{K}$ , real ou complexo. Sejam  $\|\cdot\|_X$  e  $\|\cdot\|_Y$  as normas em  $X$  e  $Y$ , respectivamente. Então*

$$\|\mathbf{T}\|_{\mathcal{L}(X,Y)} = \sup_{\substack{\mathbf{u} \in X \\ \mathbf{u} \neq \mathbf{0}}} \frac{\|\mathbf{T}\mathbf{u}\|_Y}{\|\mathbf{u}\|_X} \quad (1.5)$$

*define uma norma no espaço das transformações lineares limitadas de  $X$  em  $Y$ .*

OBSERVAÇÃO 1.8. Por conta da compacidade de conjuntos fechados limitados em espaços vetoriais de dimensão finita, o supremo em (1.5) é um máximo.

Essa classe de norma tem uma ótima propriedade.

TEOREMA 1.19. *Sejam  $X$ ,  $Y$  e  $Z$  espaços vetoriais normados sobre um mesmo corpo  $\mathbb{K}$ , real ou complexo, com normas  $\|\cdot\|_X$ ,  $\|\cdot\|_Y$  e  $\|\cdot\|_Z$ , respectivamente. Para todo  $T \in \mathcal{L}(X, Y)$  e todo  $S \in \mathcal{L}(Y, Z)$ , vale*

$$\|ST\|_{\mathcal{L}(X, Z)} \leq \|S\|_{\mathcal{L}(Y, Z)} \|T\|_{\mathcal{L}(X, Y)}.$$

OBSERVAÇÃO 1.9. Nem sempre a norma é bem comportada assim, em relação a composições. Por exemplo, considerando

$$\mathbf{A} = \mathbf{B} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

temos  $\|\mathbf{A}\|_\infty = \|\mathbf{B}\|_\infty = 1$ , de modo que o produto também é 1, no entanto

$$\mathbf{AB} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$$

e

$$\|\mathbf{AB}\|_\infty = 2 > 1 = \|\mathbf{A}\|_\infty \|\mathbf{B}\|_\infty.$$

OBSERVAÇÃO 1.10. Uma norma  $\|\cdot\|$  para transformações  $T \in \mathcal{L}(X, Y)$  entre espaços normados com normas  $\|\cdot\|_X$  e  $\|\cdot\|_Y$  é dita **consistente** quando

$$\|T\mathbf{u}\|_Y \leq \|T\| \|\mathbf{u}\|_X,$$

para todo  $\mathbf{u} \in X$ . A norma  $\|\cdot\|_{\mathcal{L}(X, Y)}$  definida por (1.5) é, naturalmente, consistente. Normas consistentes satisfazem  $\|TS\| \leq \|T\| \|S\|$ , como em (1.19).

Em particular, podemos estimar potências de um operador via

$$\|T^n\|_{\mathcal{L}(X)} \leq \|T\|_{\mathcal{L}(X)}^n.$$

OBSERVAÇÃO 1.11 (Teorema de Gelfand). Uma norma de operadores pode ser usada para se estimar o raio espectral (veja Definição 1.4) de um operador  $T \in \mathcal{L}(X)$  em um espaço normado  $X$ , graças ao Teorema de Gelfand, que diz que, para uma norma qualquer  $\|\cdot\|$  em  $\mathcal{L}(X)$ ,

$$\rho(T) = \lim_{k \rightarrow \infty} \|T^k\|^{1/k}.$$

No caso de uma norma consistente, temos sempre  $\rho(T) \leq \|T\|^{1/k}$ , ou seja, a convergência “por cima”.

OBSERVAÇÃO 1.12 (Exponencial de operadores). Vimos, na Seção 5.1, a exponencial de matrizes através da forma normal de Jordan. Usando o conceito de norma, a definição, pelo menos, de exponencial de um operador (limitado) qualquer  $T \in \mathcal{L}(X)$

em um espaço normado  $X$ , com uma norma consistente  $\|\cdot\|$  em  $\mathcal{L}(X)$ , é mais simples, visto que

$$\left\| \sum_{j=0}^m \frac{1}{j!} T^j \right\| \leq \sum_{j=0}^m \frac{1}{j!} \|T^j\| \leq \sum_{j=0}^m \frac{1}{j!} \|T\|^j \leq e^{\|T\|} < \infty,$$

de modo que a série é absolutamente convergente. Veja [Seção 5.1](#).

Caso a norma seja menor do que um, uma perturbação da identidade pelo operador continua sendo invertível.

**TEOREMA 1.20.** *Seja  $R \in \mathcal{L}(X)$  um operador em um espaço normado  $X$  real ou complexo, com norma  $\|R\|_{\mathcal{L}(X)} < 1$ . Então  $I - R$  é invertível, com inversa dada por*

$$(I - R)^{-1} = \sum_{j=0}^{\infty} R^j,$$

com

$$\|(I - R)^{-1}\|_{\mathcal{L}(X)} \leq \frac{1}{1 - \|R\|_{\mathcal{L}(X)}}$$

**DEMONSTRAÇÃO.** Observe que

$$\left\| \sum_{j=m}^{\infty} R^j \right\|_{\mathcal{L}(X)} \leq \sum_{j=m}^{\infty} \|R\|_{\mathcal{L}(X)}^j = \frac{\|R\|_{\mathcal{L}(X)}^m}{1 - \|R\|_{\mathcal{L}(X)}} \rightarrow 0, \quad m \rightarrow \infty.$$

Dessa forma, a série converge absolutamente e define um operador linear em  $X$ . Agora, note que

$$(I - R) \sum_{j=0}^m R^j = I - R^{m+1}.$$

Fazendo  $m \rightarrow \infty$ , o lado direito converge para o operador identidade e obtemos

$$(I - R) \sum_{j=0}^{\infty} R^j = I,$$

ou seja,  $I - R$  é invertível, com

$$(I - R)^{-1} = \sum_{j=0}^{\infty} R^j.$$

□

## 2. Produto interno real

O plano e o espaço têm uma estrutura geométrica a mais que diz respeito aos ângulos gerados por retas concorrentes. Em particular, temos o Teorema de Pitágoras, que nos dá uma relação direta entre os lados de um triângulo retângulo, associado à noção de ortogonalidade. Em relação à geometria analítica, isso é concretizado com a noção de produto escalar. Essa noção de produto escalar pode ser generalizada e enriquecer o estudo de espaços e operadores lineares que possuam alguma simetria em relação a esse produto interno. Vamos começar definindo essa noção abstrata de produto interno e em seguida explorando as suas propriedades e vendo como isso nos permite estudar operadores lineares com certas simetrias.

**2.1. Definição e exemplos.** No caso de espaços vetoriais reais, o produto interno toma a seguinte forma.

DEFINIÇÃO 2.1 (Produto interno real). *Um **produto interno** em espaço vetorial real  $X$  é uma forma bilinear  $((\cdot, \cdot)) : X \times X \rightarrow \mathbb{R}$  simétrica e positiva definida, i.e.*

- (i)  $((\mathbf{u} + \lambda\mathbf{v}, \mathbf{w})) = ((\mathbf{u}, \mathbf{w})) + \lambda((\mathbf{v}, \mathbf{w}))$  e  $((\mathbf{u}, \mathbf{v} + \lambda\mathbf{w})) = ((\mathbf{u}, \mathbf{v})) + \lambda((\mathbf{u}, \mathbf{w}))$ , para todo  $\mathbf{u}, \mathbf{v}, \mathbf{w} \in X$  e todo  $\lambda \in \mathbb{R}$ ;
- (ii)  $((\mathbf{u}, \mathbf{v})) = ((\mathbf{v}, \mathbf{u}))$ , para todo  $\mathbf{u}, \mathbf{v} \in X$ ; e
- (iii)  $((\mathbf{u}, \mathbf{u})) \geq 0$ , para todo  $\mathbf{u} \in X$ , com  $((\mathbf{u}, \mathbf{u})) = 0$  se, e somente se,  $\mathbf{u} = \mathbf{0}$ .

EXEMPLO 2.1 (Produto escalar em  $\mathbb{R}^n$ ). O exemplo clássico de produto interno, que é, de fato, a inspiração para o conceito mais geral de produto interno, é o *produto escalar*, definido em  $\mathbb{R}^n$  por

$$\mathbf{u} \cdot \mathbf{v} = x_1y_1 + \cdots + x_ny_n = \sum_{j=1}^n x_jy_j,$$

para todo  $\mathbf{u} = (x_1, \dots, x_n), \mathbf{v} = (y_1, \dots, y_n) \in \mathbb{R}^n$ . Uma propriedade geométrica fundamental do produto escalar é a identidade

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos \theta,$$

onde

$$\|\mathbf{u}\| = \sqrt{x_1^2 + \cdots + x_n^2}$$

é a *norma Euclidiana*, em  $\mathbb{R}^n$ , e  $\theta$  é o menor ângulo entre os vetores  $\mathbf{u}$  e  $\mathbf{v}$  (i.e. o menor ângulo entre as retas geradas por esses dois vetores). Em particular,  $\mathbf{u}$  e  $\mathbf{v}$  são ortogonais se, e somente se,  $\mathbf{u} \cdot \mathbf{v} = 0$ . O produto interno nos permite estender a noção de ângulo e de ortogonalidade para espaços e objetos mais gerais.

EXEMPLO 2.2. Dados  $a, b > 0$ , a forma bilinear

$$((\mathbf{u}, \mathbf{v})) = ax_1y_1 + bx_2y_2,$$

definida para  $\mathbf{u} = (x_1, x_2)$ ,  $\mathbf{v} = (y_1, y_2)$  em  $\mathbb{R}^2$ , nos dá um produto interno em  $\mathbb{R}^2$ . Os conjuntos

$$S_r = \{\mathbf{u} = (x, y) \in \mathbb{R}^2; \|\mathbf{u}\| = ax^2 + by^2 = r\},$$

para  $r > 0$  dado, são elipses no plano, com semi-eixos coincidindo com os eixos canônicos. Mais geralmente,

$$((\mathbf{u}, \mathbf{v})) = ax_1y_1 + bx_1y_2 + bx_2y_1 + cx_2y_2$$

nos dá uma forma bilinear simétrica em  $\mathbb{R}^2$ , que é positiva definida se, e somente se,  $a, c > 0$  e  $b^2 < ac$ , caso em que a superfície  $z = ax^2 + 2bxy + cy^2$  definida no espaço  $xyz$  forma um parabolóide com concavidade para cima. Os conjuntos

$$S_r = \{\mathbf{u} = (x, y) \in \mathbb{R}^2; \|\mathbf{u}\| = ax^2 + 2bxy + cy^2 = r\}$$

também são elipses no plano, com eixos principais podendo ser diferentes dos eixos canônicos.

EXEMPLO 2.3 (Produto interno em  $\ell^2(\mathbb{R})$ ). No espaço das seqüências reais de quadrado somável, i.e.

$$\ell^2(\mathbb{R}) = \{x = (x_n)_{n \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}; \sum_{n \in \mathbb{N}} |x_n|^2 < \infty\},$$

podemos definir o produto interno

$$((x, y)) = \sum_{n \in \mathbb{N}} x_n y_n.$$

De fato, como  $ab \leq a^2 + b^2$ , para todo  $a, b \in \mathbb{R}$ , temos

$$((x, y)) = \sum_{n \in \mathbb{N}} x_n y_n \leq \sum_{n \in \mathbb{N}} (x_n^2 + y_n^2) < \infty.$$

É fácil verificar que  $((\cdot, \cdot))$  satisfaz as hipóteses de um produto interno.

EXEMPLO 2.4 (Produto interno em  $\ell^2(\mathbb{R})$  com peso). Dada uma seqüência real positiva qualquer  $\varphi = (a_n)_{n \in \mathbb{N}}$ ,  $a_n > 0$ ,  $\forall n \in \mathbb{N}$ , podemos considerar, também, o espaço  $\ell_\varphi^2$ , i.e.  $\ell^2$  com peso  $\varphi$ , como

$$\ell_\varphi^2(\mathbb{R}) = \{x = (x_n)_{n \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}; \sum_{n \in \mathbb{N}} |x_n|^2 a_n < \infty\}.$$

A forma bilinear

$$((x, y)) = \sum_{n \in \mathbb{N}} x_n y_n a_n.$$

está bem definida e é um produto interno em  $\ell_\varphi^2$ , associado à norma

$$\|x\|^2 = \sum_{n \in \mathbb{N}} |x_n|^2 a_n$$

EXEMPLO 2.5 (Produto interno em  $L^2$ ). Dado um domínio  $\Omega \subset \mathbb{R}^n$ , o espaço  $L^2(\Omega)$  é o conjunto das funções reais  $u$  definidas em  $\Omega$  e Lebesgue mensuráveis e de quadrado integrável, i.e.

$$\|u\| = \|u\|_{L^2(\Omega)} = \int_{\Omega} |u(\mathbf{x})|^2 \, d\mathbf{x} < \infty.$$

Para duas funções  $u, v \in L^2(\Omega)$ , definimos o *produto interno em  $L^2(\Omega)$*  por

$$((u, v)) = ((u, v))_{L^2(\Omega)} = \int_{\Omega} u(\mathbf{x})v(\mathbf{x}) \, d\mathbf{x}.$$

EXEMPLO 2.6 (Produto interno em  $L^2$  com peso). Dado um domínio  $\Omega \subset \mathbb{R}^n$  e uma função contínua (ou apenas mensurável)  $\varphi$  em  $\Omega$  e positiva quase sempre, definimos o espaço  $L_\varphi^2(\Omega)$ , dito  $L^2$  com peso  $\varphi$ , como sendo o conjunto das funções reais  $u$  Lebesgue mensuráveis em  $\Omega$  e satisfazendo

$$\|u\|_\varphi = \|u\|_{L_\varphi^2(\Omega)} = \int_{\Omega} |u(\mathbf{x})|^2 \varphi(\mathbf{x}) \, d\mathbf{x} < \infty.$$

Para duas funções  $u, v \in L_\varphi^2(\Omega)$ , definimos o *produto interno em  $L_\varphi^2(\Omega)$*  por

$$((u, v))_\varphi = ((u, v))_{L_\varphi^2(\Omega)} = \int_{\Omega} u(\mathbf{x})v(\mathbf{x})\varphi(\mathbf{x}) \, d\mathbf{x}.$$

Por exemplo, considerando  $\varphi(x) = e^{-x^2}$ , os polinômios podem ser vistos como elementos de  $L_\varphi^2(\mathbb{R})$ .

**2.2. Norma proveniente do produto interno.** Graças às propriedades do produto interno, a quantidade  $\sqrt{((\mathbf{u}, \mathbf{u}))}$ , derivada do produto interno, tem as propriedades de uma norma e por isso é chamada de norma associada ao produto interno.

DEFINIÇÃO 2.2. Dado um espaço vetorial real  $X$  munido de um produto interno  $((\cdot, \cdot))$ , como  $((\mathbf{u}, \mathbf{u}))$  é não negativo, para todo  $\mathbf{u} \in X$ , definimos a norma de um vetor  $\mathbf{u}$  por

$$\|\mathbf{u}\| = \sqrt{((\mathbf{u}, \mathbf{u}))}.$$

Esta norma é chamada de norma associada ao produto interno.

TEOREMA 2.1. *Dado um espaço vetorial real  $X$  munido de um produto interno  $((\cdot, \cdot))$  e norma associada  $\|\cdot\|$ , vale a identidade*

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + 2((\mathbf{u}, \mathbf{v})) + \|\mathbf{v}\|^2. \quad (2.1)$$

DEMONSTRAÇÃO. Basta usar a bilinearidade e a simetria do produto interno, para obter

$$\begin{aligned} \|\mathbf{u} + \mathbf{v}\|^2 &= ((\mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v})) = ((\mathbf{u}, \mathbf{u})) + ((\mathbf{v}, \mathbf{u})) + ((\mathbf{u}, \mathbf{v})) + ((\mathbf{v}, \mathbf{v})) \\ &= \|\mathbf{u}\|^2 + 2((\mathbf{u}, \mathbf{v})) + \|\mathbf{v}\|^2. \end{aligned}$$

□

TEOREMA 2.2 (Desigualdade de Cauchy-Schwarz). *Dado um espaço vetorial real  $X$  munido de um produto interno  $((\cdot, \cdot))$ , com norma associada  $\|\cdot\|$ , vale a desigualdade*

$$|((\mathbf{u}, \mathbf{v}))| \leq \|\mathbf{u}\| \|\mathbf{v}\|, \quad \forall \mathbf{u}, \mathbf{v} \in X. \quad (2.2)$$

DEMONSTRAÇÃO. Dados  $\mathbf{u}, \mathbf{v} \in X$  e considerando  $s \in \mathbb{R}$  arbitrário, temos, pela positividade do produto interno, que

$$\|\mathbf{u} + s\mathbf{v}\| \geq 0,$$

para todo  $s$ . Usando a identidade (2.1), temos

$$0 \leq \|\mathbf{u} + s\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + 2s((\mathbf{u}, \mathbf{v})) + s^2\|\mathbf{v}\|^2.$$

Se  $\mathbf{v} = \mathbf{0}$ , então devemos ter  $((\mathbf{u}, \mathbf{v})) = 0$ , caso contrário o lado direito seria negativo para valores suficientemente grandes em módulo e com sinal contrário ao do produto interno. Nesse caso, a desigualdade é trivialmente satisfeita.

Para  $\mathbf{v} \neq \mathbf{0}$ , o lado direito é um polinômio do segundo grau e esse polinômio deve ser sempre não negativo. Em particular, o seu valor mínimo deve ser não negativo. O ponto de mínimo ocorre em

$$s = -\frac{((\mathbf{u}, \mathbf{v}))}{\|\mathbf{v}\|^2}.$$

Nesse ponto, deduzimos que

$$0 \leq \|\mathbf{u}\|^2 - 2\frac{((\mathbf{u}, \mathbf{v}))}{\|\mathbf{v}\|^2}((\mathbf{u}, \mathbf{v})) + \frac{((\mathbf{u}, \mathbf{v}))^2}{\|\mathbf{v}\|^4}\|\mathbf{v}\|^2 = \|\mathbf{u}\|^2 - \frac{((\mathbf{u}, \mathbf{v}))^2}{\|\mathbf{v}\|^2}.$$

Multiplicando por  $\|\mathbf{v}\|^2$ , obtemos

$$((\mathbf{u}, \mathbf{v}))^2 \leq \|\mathbf{u}\|^2 \|\mathbf{v}\|^2.$$

Extraindo a raiz quadrada, obtemos a desigualdade desejada. □

TEOREMA 2.3 (Desigualdade triangular). *Dado um espaço vetorial real  $X$  munido de um produto interno  $((\cdot, \cdot))$ , a norma  $\|\cdot\|$  associada ao produto interno satisfaz a desigualdade triangular*

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|, \quad \forall \mathbf{u}, \mathbf{v} \in X. \quad (2.3)$$

DEMONSTRAÇÃO. Usando a identidade (2.1) do Teorema 2.1 e a desigualdade de Cauchy-Schwartz (2.2), deduzimos que

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + 2((\mathbf{u}, \mathbf{v})) + \|\mathbf{v}\|^2 \leq \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\|\mathbf{v}\| + \|\mathbf{v}\|^2 = (\|\mathbf{u}\| + \|\mathbf{v}\|)^2.$$

Extraindo a raiz quadrada, obtemos a desigualdade desejada.  $\square$

OBSERVAÇÃO 2.1. Assim, a quantidade  $\|\cdot\|$  é um funcional de  $X$  em  $\mathbb{R}$  com as seguintes propriedades de norma:

- (i) (positiva definida)  $\|\mathbf{u}\| \geq 0$ , para todo  $\mathbf{u} \in X$ , com  $\|\mathbf{u}\| = 0$  se, e somente se,  $\mathbf{u} = \mathbf{0}$ ;
- (ii) (homogênea)  $\|\lambda\mathbf{u}\| = |\lambda|\|\mathbf{u}\|$ , para todo  $\mathbf{u} \in X$  e todo  $\lambda \in \mathbb{R}$ ; e
- (iii) (desigualdade triangular)  $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$ , para todo  $\mathbf{u}, \mathbf{v} \in X$ .

Com essa norma, podemos definir uma topologia em  $X$  e considerar a continuidade de uma função  $f : X \rightarrow \mathbb{R}$  ou de uma função  $f : X \rightarrow X$  ou, mais geralmente, de uma função  $f : X \rightarrow Y$  entre dois espaços vetoriais reais  $X$  e  $Y$  com produto interno. Em particular, a própria função norma  $f(\mathbf{u}) = \|\mathbf{u}\|$  é contínua de  $X$  em  $\mathbb{R}$ , como estabelecido a seguir.

OBSERVAÇÃO 2.2. Já vimos, no Teorema 1.1, que toda norma é contínua, o que inclui as normas proveniente de um produto interno. No caso de uma norma proveniente de um produto interno, temos uma regularidade maior, de diferenciabilidade da *norma quadrática*, como é conhecido o quadrado da norma (veja Observação 2.3).

OBSERVAÇÃO 2.3 (Diferenciabilidade da norma quadrática). Dado um espaço vetorial real  $X$  munido de um produto interno  $((\cdot, \cdot))$ , com norma associada  $\|\cdot\|$ . Então a *norma quadrática*  $\|\cdot\|^2 : X \rightarrow \mathbb{R}$  é diferenciável, com diferencial, em um ponto  $\mathbf{u}$ , dada por  $2\mathbf{u}$ , no sentido de que

$$\frac{1}{\|\mathbf{h}\|} \left| \|\mathbf{u} + \mathbf{h}\|^2 - \|\mathbf{u}\|^2 - 2((\mathbf{u}, \mathbf{h})) \right| = \|\mathbf{h}\| \rightarrow 0,$$

quando  $\mathbf{h} \rightarrow \mathbf{0}$ , com  $\mathbf{h} \neq \mathbf{0}$ . Em particular, vale a derivada direcional

$$\lim_{s \neq 0, s \rightarrow 0} \frac{\|\mathbf{u} + s\mathbf{v}\|^2 - \|\mathbf{u}\|^2}{|s|} = 2((\mathbf{u}, \mathbf{v})).$$

Para  $\mathbf{u} \in X$  fixo, a função  $\mathbf{h} \mapsto ((2\mathbf{u}, \mathbf{h}))$  é um funcional linear em  $X$ , ou seja, é um elemento  $\mathbf{f}$  de  $X^*$ , cuja ação sobre um vetor  $\mathbf{u} \in X$  é representada pelo produto

interno com  $2\mathbf{u}$ , i.e.  $\langle \mathbf{f}, \mathbf{h} \rangle = ((2\mathbf{u}, \mathbf{h}))$ . Isso, na verdade, faz parte de um resultado importante de representação de funcionais lineares em espaços com produto interno, conhecido como Teorema da Representação de Riesz, que veremos em seguida.

### 3. Ortogonalidade em espaços vetoriais reais

#### 3.1. Vetores e conjuntos ortogonais. Motivados pela identidade

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos \theta,$$

válida para o produto escalar em  $\mathbb{R}^n$ , onde  $\|\cdot\|$  é a norma Euclidiana e  $\theta$  é o menor ângulo entre os vetores  $\mathbf{u}$  e  $\mathbf{v}$ , podemos definir, mais geralmente, o ângulo entre vetores em um espaço vetorial qualquer com produto interno, via

$$\cos \theta = \frac{((\mathbf{u}, \mathbf{v}))}{\|\mathbf{u}\| \|\mathbf{v}\|},$$

onde, agora,  $\|\cdot\|$  é a norma associada a um produto interno qualquer. Graças à desigualdade de Cauchy-Schwarz, o lado direito acima é um número real entre  $-1$  e  $1$  e, portanto, define unicamente um ângulo no intervalo  $0 \leq \theta \leq \pi$ , sendo interpretado como o *menor* ângulo entre os subespaços gerados por  $\mathbf{u}$  e  $\mathbf{v}$ . Quando  $((\mathbf{u}, \mathbf{v})) = 0$ , temos  $\theta = \pi/2$  e os vetores e os seus subespaços gerados são ortogonais entre si.

**DEFINIÇÃO 3.1** (Vetores ortogonais). *Seja  $X$  um espaço vetorial real com produto interno  $((\cdot, \cdot))$ . Dois vetores  $\mathbf{u}, \mathbf{v}$  em um espaço vetorial real  $X$  munido de um produto interno  $((\cdot, \cdot))$  são ditos **ortogonais** em relação a esse produto interno, quando*

$$((\mathbf{u}, \mathbf{v})) = 0.$$

Nesse caso, escrevemos  $\mathbf{u} \perp \mathbf{v}$ .

**EXEMPLO 3.1.** Os vetores  $(1, 1)$  e  $(-1, 1)$  são ortogonais em  $\mathbb{R}^2$ , em relação ao produto escalar, visto que

$$(1, 1) \cdot (-1, 1) = -1 + 1 = 0.$$

Já em relação ao produto interno

$$((\mathbf{u}, \mathbf{v})) = 2x_1y_1 + 3x_2y_2,$$

definido para  $\mathbf{u} = (x_1, x_2)$  e  $\mathbf{v} = (y_1, y_2)$ , esses vetores não são ortogonais, enquanto que os vetores  $(1, 2)$  e  $(-3, 1)$  são ortogonais entre si, visto que

$$(((1, 2), (-3, 1))) = 2(1)(-3) + 3(2)(1) = -6 + 6 = 0.$$

Observe que os conjuntos  $\|\mathbf{u}\|^2 = ((\mathbf{u}, \mathbf{u})) = c$ ,  $c > 0$ , em relação à norma desse produto interno, são elipses no plano cartesiano.

EXEMPLO 3.2. Os polinômios  $x$  e  $x^2$  são ortogonais no espaço  $L^2(-1, 1)$  com o produto interno canônico de  $L^2$ , visto que

$$((x, x^2))_{L^2(-1,1)} = \int_{-1}^1 xx^2 dx = \int_{-1}^1 x^3 dx = \frac{x^4}{4} \Big|_{x=-1}^{x=1} = \frac{1}{4} - \frac{1}{4} = 0.$$

Já em  $L^2(0, 1)$ , eles não são ortogonais, com

$$((x, x^2))_{L^2(0,1)} = \int_0^1 xx^2 dx = \frac{x^4}{4} \Big|_{x=0}^{x=1} = \frac{1}{4}.$$

Esse conceito de ortogonalidade pode ser estendido a conjuntos de vetores, como se segue.

DEFINIÇÃO 3.2 (Conjuntos ortogonais entre si). *Seja  $X$  um espaço vetorial real com produto interno  $((\cdot, \cdot))$ . Um vetor  $\mathbf{u}$  é dito ortogonal a um subconjunto  $A \subset X$  quando  $\mathbf{u} \perp \mathbf{v}$  para todo  $\mathbf{v} \in A$ . Nesse caso, escrevemos  $\mathbf{u} \perp A$ . Mais geralmente, dois subconjuntos  $A$  e  $B$  de  $X$  são ditos ortogonais entre si quando  $\mathbf{u} \perp \mathbf{v}$  para quaisquer  $\mathbf{u} \in A$  e  $\mathbf{v} \in B$ . Nesse caso, escrevemos  $A \perp B$ .*

EXEMPLO 3.3. O vetor  $(1, 1, 1)$  é ortogonal ao plano  $x + y + z = 0$ , no espaço cartesiano  $xyz$ , visto que

$$(1, 1, 1) \cdot (x, y, z) = x + y + z = 0,$$

para todo  $(x, y, z)$  nesse plano.

A partir de um dado conjunto, podemos, ainda, considerar o subconjunto de todos os vetores ortogonais ao conjunto dado.

DEFINIÇÃO 3.3 (Subespaço ortogonal a um conjunto). *Seja  $X$  um espaço vetorial real com produto interno  $((\cdot, \cdot))$ . Dado um subconjunto  $A \subset X$ , denotamos o seu perpendicular por*

$$A^\perp = \{\mathbf{u} \in X; \mathbf{u} \perp A\}.$$

EXEMPLO 3.4. Para encontrar todos os vetores  $(a, b, c)$  ortogonais ao plano

$$P = \{(x, y, z) \in \mathbb{R}^3; x + y + z = 0\},$$

temos a condição

$$(a, b, c) \cdot (x, y, z) = ax + by + cz = 0,$$

para todo  $(x, y, z)$  nesse plano. Escolhendo  $y = 0$  e  $x \in \mathbb{R}$  arbitrário, temos  $z = -x$  e

$$0 = ax + by + cz = ax + 0 - cx = (a - c)x,$$

para todo  $x$ , logo  $c = a$ . Da mesma forma, escolhendo  $z = 0$  e  $x$  arbitrário, temos  $y = -x$  e

$$0 = ax + by + cz = ax - bx + 0 = (a - b)x,$$

para todo  $x$ , logo  $b = a$ . Todo vetor da forma  $(a, a, a)$ ,  $a \in \mathbb{R}$  é ortogonal ao plano, já que

$$(a, a, a) \cdot (x, y, z) = ax + ay + az = a(x + y + z) = 0,$$

para todo  $(x, y, z)$  no plano. Logo, o ortogonal ao plano é, exatamente, a reta

$$P^\perp = \{(a, a, a); a \in \mathbb{R}\} = \text{span}\{(1, 1, 1)\}.$$

Independentemente de um subconjunto  $A \subset X$  ser subespaço ou não, o seu perpendicular  $A^\perp$  é sempre um subespaço linear.

**TEOREMA 3.1.** *Seja  $X$  um espaço vetorial real com produto interno  $(\cdot, \cdot)$ . Dado um subconjunto  $A \subset X$ , o seu perpendicular  $A^\perp$  é um subespaço vetorial de  $X$ . Se  $S$  é um subespaço vetorial de  $X$ , então  $S$  e  $S^\perp$  são linearmente independentes.*

**EXEMPLO 3.5.** Considere o segmento de reta

$$A = \{(x, y) \in \mathbb{R}^2; y = x, 0 \leq x \leq 1\} = \{s(1, 1); 0 \leq s \leq 1\}.$$

O seu ortogonal é o subespaço

$$A^\perp = \{(x, y) \in \mathbb{R}^2; y = -x\} = \text{span}\{(-1, 1)\}.$$

Podemos, também, considerar o perpendicular do perpendicular, que também é sempre um subespaço vetorial:

$$A^{\perp\perp} = \{(x, y) \in \mathbb{R}^2; y = x\} = \text{span}\{(1, 1)\}.$$

No caso em que  $X$  é de dimensão finita e  $S$  é um subespaço linear de  $X$ , vale, ainda, conforme será visto no **Teorema 3.4**, que  $X = S \oplus S^\perp$ , mas antes precisamos falar de base ortonormal e projeção ortogonal.

**3.2. Projeção ortogonal e o processo de Gram-Schmidt.** Um conceito fundamental associado ao produto interno é o de projeção ortogonal.

**DEFINIÇÃO 3.4 (Projeção ortogonal).** *Seja  $X$  um espaço vetorial real com produto interno  $(\cdot, \cdot)$ . Uma projeção  $P : X \rightarrow X$  sobre um subespaço  $S$  é dita uma **projeção ortogonal** quando  $\ker P = S^\perp$ , ou seja, quando  $P$  é uma projeção sobre  $S$  ao longo do subespaço perpendicular  $S^\perp$ .*

OBSERVAÇÃO 3.1. Implícito na **Definição 3.4** está o fato de que  $X = S \oplus S^\perp$ , que ainda não provamos, mas que é verdade para todo subespaço  $S$ , quando  $X$  é de dimensão finita. Veremos, junto com isso, que uma tal projeção ortogonal sempre existe, em dimensão finita, ou, pelo menos, quando  $S$  é de dimensão finita.

Vamos começar mostrando a existência da projeção ortogonal sobre um subespaço unidimensional.

TEOREMA 3.2 (Projeção ortogonal na direção de um vetor). *Seja  $X$  um espaço vetorial real com produto interno  $((\cdot, \cdot))$  e norma associada  $\|\cdot\|$ . Seja  $\mathbf{v} \in X$  um vetor não nulo. Então*

$$P_{\mathbf{v}}\mathbf{u} = \frac{((\mathbf{u}, \mathbf{v}))}{\|\mathbf{v}\|^2}\mathbf{v}$$

define um operador de projeção  $P_{\mathbf{v}} : X \rightarrow X$  sobre o subespaço  $S = \text{span}\{\mathbf{v}\}$  gerado por  $\mathbf{v}$ , ao longo do subespaço ortogonal  $S^\perp$ .

DEMONSTRAÇÃO. Como o produto interno é bilinear, sendo, em particular, linear no primeiro argumento, segue que  $\mathbf{u} \mapsto P_{\mathbf{v}}\mathbf{u}$  é, de fato, um operador linear em  $\mathbf{u}$ . Para ver que é uma projeção, observe que

$$P_{\mathbf{v}}^2\mathbf{u} = P_{\mathbf{v}}\left(\frac{((\mathbf{u}, \mathbf{v}))}{\|\mathbf{v}\|^2}\mathbf{v}\right) = \frac{((\mathbf{u}, \mathbf{v}))}{\|\mathbf{v}\|^2}P_{\mathbf{v}}\mathbf{v}.$$

Como

$$P_{\mathbf{v}}\mathbf{v} = \frac{((\mathbf{v}, \mathbf{v}))}{\|\mathbf{v}\|^2}\mathbf{v} = \mathbf{v},$$

então,

$$P_{\mathbf{v}}^2\mathbf{u} = \frac{((\mathbf{u}, \mathbf{v}))}{\|\mathbf{v}\|^2}\mathbf{v} = P_{\mathbf{v}}\mathbf{u},$$

para todo  $\mathbf{u} \in X$ , mostrando que  $P^2 = P$  e, portanto,  $P$  é uma projeção em  $X$ .

Obviamente,  $P\mathbf{u}$  é um múltiplo de  $\mathbf{v}$ . Além disso,

$$P(\alpha\mathbf{v}) = \alpha P\mathbf{v} = \alpha\mathbf{v},$$

para todo  $\alpha \in \mathbb{R}$ , ou seja,

$$\text{Im } P = \text{span}\{\mathbf{v}\},$$

mostrando que  $P$  é uma projeção sobre  $\text{span}\{\mathbf{v}\}$ .

Resta mostrar que  $\ker P = \text{span}\{\mathbf{v}\}^\perp$ . Para isso, basta observar que

$$\mathbf{u} \in \text{span}\{\mathbf{v}\}^\perp \Leftrightarrow ((\mathbf{u}, \mathbf{v})) = 0 \Leftrightarrow P_{\mathbf{v}}\mathbf{u} = \frac{((\mathbf{u}, \mathbf{v}))}{\|\mathbf{v}\|^2}\mathbf{v} = \mathbf{0} \Leftrightarrow \mathbf{u} \in \ker P,$$

concluindo a demonstração. □

Podemos estender a construção acima a um subespaço vetorial  $S$  de dimensão finita. Para isso, usamos um processo de ortogonalização conhecido como processo de Gram-Schmidt.

**TEOREMA 3.3** (Processo de ortogonalização de Gram-Schmidt). *Seja  $X$  um espaço vetorial real com produto interno  $((\cdot, \cdot))$  e norma associada  $\|\cdot\|$ . Dada uma sequência  $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$  de vetores linearmente independentes, a sequência  $\{\mathbf{w}_1, \dots, \mathbf{w}_m\}$  dada por*

$$\mathbf{w}_j = \frac{\mathbf{w}'_j}{\|\mathbf{w}'_j\|}, \quad \mathbf{w}'_j = \mathbf{v}_j - P_{\mathbf{w}_1} \mathbf{v}_j - \dots - P_{\mathbf{w}_{j-1}} \mathbf{v}_j,$$

para  $j = 1, \dots, m$ , define vetores ortogonais entre si, com norma unitária e gerando os mesmos subespaços, i.e.

$$\|\mathbf{w}_j\| = 1, \quad ((\mathbf{w}_i, \mathbf{w}_j)) = 0, \quad \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_j\} = \text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_j\},$$

para todo  $i, j = 1, \dots, m$ , com  $i \neq j$ .

**DEMONSTRAÇÃO.** Mais explicitamente, temos

$$\begin{aligned} \mathbf{w}_1 &= \frac{\mathbf{v}'_1}{\|\mathbf{v}'_1\|}, & \mathbf{v}'_1 &= \mathbf{v}_1, \\ \mathbf{w}_2 &= \frac{\mathbf{w}'_2}{\|\mathbf{w}'_2\|}, & \mathbf{w}'_2 &= \mathbf{v}_2 - P_{\mathbf{w}_1} \mathbf{v}_2, \\ & \vdots & & \vdots \\ \mathbf{w}_j &= \frac{\mathbf{w}'_j}{\|\mathbf{w}'_j\|}, & \mathbf{w}'_j &= \mathbf{v}_j - P_{\mathbf{w}_1} \mathbf{v}_j - \dots - P_{\mathbf{w}_{j-1}} \mathbf{v}_j. \end{aligned}$$

Observe que cada  $\mathbf{w}_j$  está normalizado, visto que

$$\|\mathbf{w}_j\| = \left\| \frac{\mathbf{w}'_j}{\|\mathbf{w}'_j\|} \right\| = \frac{1}{\|\mathbf{w}'_j\|} \|\mathbf{w}'_j\| = 1.$$

Dessa forma, a projeção ortogonal sobre cada  $\mathbf{w}_i$  se reduz a

$$P_{\mathbf{w}_i} \mathbf{u} = \frac{((\mathbf{u}, \mathbf{w}_i))}{\|\mathbf{w}_i\|^2} \mathbf{w}_i = ((\mathbf{u}, \mathbf{w}_i)) \mathbf{w}_i.$$

Por construção, cada  $\mathbf{w}_j$  é uma combinação linear dos vetores  $\mathbf{v}_1, \dots, \mathbf{v}_j$ , portanto é imediato deduzir que

$$\text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_j\} = \text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_j\},$$

para todo  $j = 1, \dots, m$ .

Resta mostrar a ortogonalidade entre vetores diferentes. Vamos provar isso por indução. Começamos mostrando que  $\mathbf{w}_1 \perp \mathbf{w}_2$ . Nesse caso, temos

$$\begin{aligned}
 ((\mathbf{w}_2, \mathbf{w}_1)) &= \frac{1}{\|\mathbf{w}'_2\| \|\mathbf{v}_1\|} ((\mathbf{w}'_2, \mathbf{v}_1)) \\
 &= \frac{1}{\|\mathbf{w}'_2\| \|\mathbf{v}_1\|} (((\mathbf{v}_2, \mathbf{v}_1)) - ((\mathbf{P}_{\mathbf{w}_1} \mathbf{v}_2, \mathbf{v}_1))) \\
 &= \frac{1}{\|\mathbf{w}'_2\| \|\mathbf{v}_1\|} (((\mathbf{v}_2, \mathbf{v}_1)) - ((\mathbf{v}_2, \mathbf{w}_1))((\mathbf{w}_1, \mathbf{v}_1))) \\
 &= \frac{1}{\|\mathbf{w}'_2\| \|\mathbf{v}_1\|} \left( ((\mathbf{v}_2, \mathbf{v}_1)) - \frac{((\mathbf{v}_2, \mathbf{v}_1))}{\|\mathbf{v}_1\|^2} ((\mathbf{v}_1, \mathbf{v}_1)) \right) \\
 &= \frac{1}{\|\mathbf{w}'_2\| \|\mathbf{v}_1\|} (((\mathbf{v}_2, \mathbf{v}_1)) - ((\mathbf{v}_2, \mathbf{v}_1))) \\
 &= 0.
 \end{aligned}$$

Portanto,  $\mathbf{w}_2 \perp \mathbf{w}_1$ . Agora, prosseguindo por indução, assumimos que  $\mathbf{w}_l \perp \mathbf{w}_i$  para  $i, l = 1, \dots, j-1$ ,  $i \neq l$ , onde  $j \leq m$ . Vamos mostrar que  $\mathbf{w}_j \perp \mathbf{w}_i$ , para  $i = 1, \dots, j-1$ . De fato,

$$\begin{aligned}
 ((\mathbf{w}_j, \mathbf{w}_i)) &= \frac{1}{\|\mathbf{w}'_j\| \|\mathbf{w}'_i\|} ((\mathbf{w}'_j, \mathbf{w}'_i)) \\
 &= \frac{1}{\|\mathbf{w}'_j\| \|\mathbf{w}'_i\|} \left( ((\mathbf{v}_j, \mathbf{w}'_i)) - \sum_{l=1}^{j-1} ((\mathbf{P}_{\mathbf{w}_l} \mathbf{v}_j, \mathbf{w}'_i)) \right) \\
 &= \frac{1}{\|\mathbf{w}'_j\| \|\mathbf{w}'_i\|} \left( ((\mathbf{v}_j, \mathbf{w}'_i)) - \sum_{l=1}^{j-1} ((\mathbf{v}_j, \mathbf{w}_l))((\mathbf{w}_l, \mathbf{w}'_i)) \right) \\
 &= \frac{1}{\|\mathbf{w}'_j\| \|\mathbf{w}'_i\|} (((\mathbf{v}_j, \mathbf{w}'_i)) - ((\mathbf{v}_j, \mathbf{w}_i))((\mathbf{w}'_i, \mathbf{w}_i))) \\
 &= \frac{1}{\|\mathbf{w}'_j\| \|\mathbf{w}'_i\|} (((\mathbf{v}_j, \mathbf{w}'_i)) - ((\mathbf{v}_j, \mathbf{w}'_i))) \\
 &= 0.
 \end{aligned}$$

Portanto,  $\mathbf{w}_j \perp \mathbf{w}_i$ , para  $i = 1, \dots, j-1$ , onde  $j \leq m$  é arbitrário. Ou seja, a indução vale para  $j = 1, \dots, m$ . Isso completa a demonstração do processo de ortogonalização de Gram-Schmidt.  $\square$

Com base do processo de Gram-Schmidt, podemos construir a projeção ortogonal sobre subespaços vetoriais de dimensão finita.

TEOREMA 3.4. *Seja  $X$  um espaço vetorial real com produto interno  $((\cdot, \cdot))$  e seja  $S \subset X$  um subespaço vetorial de  $X$  de dimensão finita. Então existe uma projeção ortogonal  $P_S : X \rightarrow X$  sobre  $S$  e vale*

$$X = S \oplus S^\perp.$$

DEMONSTRAÇÃO. Já vimos, no Teorema 3.1, que  $S$  e  $S^\perp$  são independentes. Falta mostrar que  $S + S^\perp = X$ . Para isso, vamos construir uma projeção ortogonal sobre  $S$ .

Sendo  $S$  de dimensão finita,  $S$  possui uma base finita, a menos que  $S = \{\mathbf{0}\}$ . Nesse caso em que  $S$  é trivial, basta tomar  $P\mathbf{u} = \mathbf{0}$ , que naturalmente temos  $S^\perp = X = \ker P$  e  $\text{Im } P = S$ . A dificuldade maior está no caso em que  $S$  tem dimensão finita positiva. Seja, então,  $\mathcal{b} = \{\mathbf{v}_1, \dots, \mathbf{v}_m\}$  uma base de  $S$ , para algum  $m \in \mathbb{N}$ .

Pelo processo de Gram-Schmidt, visto no Teorema 3.3, obtemos uma nova base  $\{\mathbf{w}_1, \dots, \mathbf{w}_m\}$  para  $S$ , tal que  $\|\mathbf{w}_j\| = 1$  e  $((\mathbf{w}_i, \mathbf{w}_j)) = 0$ , para  $i \neq j$ .

Agora, definimos

$$P_S \mathbf{u} = \sum_{j=1}^m P_{\mathbf{w}_j} \mathbf{u} = \sum_{j=1}^m ((\mathbf{u}, \mathbf{w}_j)) \mathbf{w}_j.$$

Como os elementos da base  $\{\mathbf{w}_1, \dots, \mathbf{w}_m\}$  são ortogonais entre si e normalizados, é fácil ver que

$$P_S^2 \mathbf{u} = \sum_{j=1}^m ((P_S \mathbf{u}, \mathbf{w}_j)) \mathbf{w}_j = \sum_{j=1}^m ((\sum_{i=1}^m ((\mathbf{u}, \mathbf{w}_i)) \mathbf{w}_i, \mathbf{w}_j)) \mathbf{w}_j = \sum_{j=1}^m ((\mathbf{u}, \mathbf{w}_j)) \mathbf{w}_j = P_S \mathbf{u},$$

ou seja,  $P$  é, de fato, uma projeção. Falta provar que é uma projeção ortogonal. Para isso, vamos mostrar que a imagem de  $Q_S = I - P_S$  está no perpendicular  $S^\perp$  de  $S$ . Dado  $\mathbf{u} \in X$ , seja  $\mathbf{v} = Q_S \mathbf{u} = \mathbf{u} - P_S \mathbf{u}$ . Vamos mostrar que  $\mathbf{v} \perp \mathbf{w}_j$ , para todo  $j = 1, \dots, m$ . De fato, pela definição da projeção  $P_S$ ,

$$\begin{aligned} ((\mathbf{v}, \mathbf{w}_j)) &= ((\mathbf{u} - P_S \mathbf{u}, \mathbf{w}_j)) \\ &= ((\mathbf{u}, \mathbf{w}_j)) - ((\sum_{i=1}^m ((\mathbf{u}, \mathbf{w}_i)) \mathbf{w}_i, \mathbf{w}_j)) \\ &= ((\mathbf{u}, \mathbf{w}_j)) - \sum_{i=1}^m ((\mathbf{u}, \mathbf{w}_i)) ((\mathbf{w}_i, \mathbf{w}_j)). \end{aligned}$$

Usando que  $\mathbf{w}_1, \dots, \mathbf{w}_m$  são ortogonais entre si e de norma um, temos

$$\sum_{i=1}^m ((\mathbf{u}, \mathbf{w}_i)) ((\mathbf{w}_i, \mathbf{w}_j)) = ((\mathbf{u}, \mathbf{w}_j)).$$

Logo,

$$((\mathbf{v}, \mathbf{w}_j)) = ((\mathbf{u}, \mathbf{w}_j)) - ((\mathbf{u}, \mathbf{w}_j)) = 0,$$

mostrando que  $\mathbf{v} \perp S$ , ou seja,  $\mathbf{v} \in S^\perp$ .

Mostramos, então, que todo vetor  $\mathbf{u} \in X$  se escreve da forma  $\mathbf{u} = \mathbf{w} + \mathbf{v}$ , onde  $\mathbf{w} = P_S \mathbf{u} \in S$  e  $\mathbf{v} = (I - P_S) \mathbf{u} = \mathbf{u} - P_S \mathbf{u} \in S^\perp$ , de modo que

$$X = S \oplus S^\perp.$$

e  $P_S$  é a projeção ortogonal sobre  $S$ .  $\square$

**OBSERVAÇÃO 3.2.** O resultado do **Teorema 3.4** pode ser usado para demonstrar a existência de projeção ortogonal sobre subespaços de *codimensão* finita. Veja **Exercício 6.2**.

**OBSERVAÇÃO 3.3.** Como sempre, em dimensão infinita as coisas são mais delicadas. Por exemplo, o conjunto  $S = \mathcal{C}(I)$  das funções reais contínuas definidas em um intervalo  $I$  formam um subespaço do espaço com produto interno  $X = L^2(I)$ . No entanto, o seu perpendicular é o subespaço trivial,  $S^\perp = \{0\}$  e  $X$  não é soma direta dos dois, i.e.  $X \neq S \oplus S^\perp = S$ .

O Teorema de Pitágoras em espaços com produto interno continua válido no caso complexo.

**TEOREMA 3.5.** *Seja  $X$  um espaço vetorial real com produto interno  $((\cdot, \cdot))$  e suponha que  $P_S$  seja uma projeção ortogonal sobre um subespaço  $S$ . Então  $P_{S^\perp} = I - P_S$  é a projeção ortogonal sobre  $S^\perp$  e vale a identidade*

$$\|\mathbf{u}\|^2 = \|P_S \mathbf{u}\|^2 + \|P_{S^\perp} \mathbf{u}\|^2, \quad \forall \mathbf{u} \in X. \quad (3.1)$$

**DEMONSTRAÇÃO.** De fato, como  $P_S \mathbf{u} \in S$  e  $P_{S^\perp} \mathbf{u} \in S^\perp$  e os subespaços são ortogonais entre si, então

$$((P_S \mathbf{u}, P_{S^\perp} \mathbf{u})) = 0.$$

Com isso, o resultado desejada sai da identidade (2.1) aplicada a  $\mathbf{u} = P_S \mathbf{u} + P_{S^\perp} \mathbf{u}$ .  $\square$

A propriedade de ortogonalidade entre o projetor e o seu complemento pode ser usada para nos dar uma caracterização de projeção ortogonal.

**TEOREMA 3.6.** *Seja  $X$  um espaço vetorial real com produto interno  $((\cdot, \cdot))$  e suponha que  $P$  seja uma projeção em  $X$ . Então  $P$  é projeção ortogonal se, e somente se,*

$$((P\mathbf{u}, Q\mathbf{u})) = 0,$$

para todo  $\mathbf{u} \in X$ , onde  $Q = I - P$ .

DEMONSTRAÇÃO. Suponha que  $P$  seja uma projeção ortogonal sobre  $U$ , ao longo de  $V$ , de modo que  $V = U^\perp$ . Nesse caso  $P\mathbf{u} \in U$  e  $Q\mathbf{u} \in V = U^\perp$ , de modo que  $P\mathbf{u}$  e  $Q\mathbf{u}$  são ortogonais entre si e o produto interno entre eles se anula.

Suponha agora que o produto interno entre essas projeções se anula. Sejam  $\mathbf{w} \in U = \text{Im}(P)$  e  $\mathbf{v} \in V = \text{ker}(P)$  arbitrários. Então podemos escrever  $\mathbf{u} = \mathbf{w} + \mathbf{v}$ , com  $\mathbf{w} = P\mathbf{u}$  e  $\mathbf{v} = Q\mathbf{u}$ . Dessa forma,

$$((\mathbf{w}, \mathbf{v})) = ((P\mathbf{u}, Q\mathbf{v})) = 0,$$

mostrando que  $\mathbf{w} \perp \mathbf{v}$ . Como isso é válido para  $\mathbf{w}$  e  $\mathbf{v}$  arbitrários em  $U$  e  $V$ , respectivamente, então  $U \perp V$  e  $P$  é projeção sobre  $U$  ao longo do ortogonal  $V = U^\perp$  de  $U$ , ou seja,  $P$  é projeção ortogonal.  $\square$

A projeção ortogonal tem a propriedade de minimizar a distância até um subespaço vetorial, conforme expressado pelo teorema a seguir.

TEOREMA 3.7. *Seja  $X$  um espaço vetorial real com produto interno  $((\cdot, \cdot))$  e norma associada  $\|\cdot\|$ . Seja  $S \subset X$  e suponha que  $P_S$  seja a projeção ortogonal sobre  $S$ . Então*

$$\|\mathbf{u} - P_S\mathbf{u}\| \leq \|\mathbf{u} - \mathbf{v}\|, \quad \forall \mathbf{v} \in S.$$

DEMONSTRAÇÃO. De fato, usando a identidade (2.1)

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u} - P_S\mathbf{u} + P_S\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u} - P_S\mathbf{u}\|^2 + 2((\mathbf{u} - P_S\mathbf{u}, P_S\mathbf{u} - \mathbf{v})) + \|P_S\mathbf{u} - \mathbf{v}\|^2.$$

Como ambos  $P_S\mathbf{u}$  e  $\mathbf{v}$  estão em  $S$ , temos  $P_S\mathbf{u} - \mathbf{v} \in S$ , enquanto que  $\mathbf{u} - P_S\mathbf{u}$  é ortogonal a  $S$ , de modo que o produto interno se anula, nos deixando

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u} - P_S\mathbf{u}\|^2 + \|P_S\mathbf{u} - \mathbf{v}\|^2 \geq \|\mathbf{u} - P_S\mathbf{u}\|^2.$$

Extraindo a raiz quadrada, chegamos na identidade desejada.  $\square$

**3.3. Base ortonormal.** No processo de construção da projeção ortogonal  $P_S$  sobre um subespaço  $S$  de dimensão finita, começamos com uma base  $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$  de  $S$  e construímos uma outra base  $\{\mathbf{w}_1, \dots, \mathbf{w}_m\}$  com algumas propriedades especiais, a saber, que todos os vetores tem norma unitária e os vetores são ortogonais entre si. Esse tipo de base leva um nome especial.

DEFINIÇÃO 3.5. *Seja  $X$  um espaço vetorial real com produto interno  $((\cdot, \cdot))$  e norma associada  $\|\cdot\|$  e dimensão finita  $n = \dim(X) < \infty$ . Uma **base ortonormal** de  $X$  é uma base  $\mathcal{B} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  tal que cada vetor tem norma unitária e os vetores são ortogonais entre si, ou seja,*

$$\|\mathbf{w}_j\| = 1, \quad ((\mathbf{w}_i, \mathbf{w}_j)) = 0, \quad i, j = 1, \dots, n, \quad i \neq j.$$

EXEMPLO 3.6. Os vetores  $(1, 1)$  e  $(-1, 1)$  são ortogonais mas não são unitários. Mas basta normalizá-los para obter uma base ortonormal:

$$\mathcal{b} = \left\{ \frac{\sqrt{2}}{2}(1, 1), \frac{\sqrt{2}}{2}(-1, 1) \right\}.$$

O processo de ortogonalização de Gram-Schmidt nos garante a existência de uma base ortonormal de qualquer subespaço de dimensão finita.

TEOREMA 3.8. *Seja  $X$  um espaço vetorial real com produto interno e seja  $S \subset X$  um subespaço de  $X$  não trivial de dimensão finita. Então  $S$  possui uma base ortonormal. Mais precisamente, se  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  é uma base de  $S$ , para  $n = \dim(S)$ , então existe uma base ortonormal  $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  com a propriedade*

$$\text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_j\} = \text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_j\},$$

para todo  $j = 1, \dots, n$ .

DEMONSTRAÇÃO. Basta aplicar o processo de ortogonalização de Gram-Schmidt, descrito no Teorema 3.3.  $\square$

EXEMPLO 3.7. A base canônica  $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ , onde  $\mathbf{e}_j = (\delta_{ij})_{i=1, \dots, n}$ , é um exemplo de base ortonormal em relação ao produto escalar de  $\mathbb{R}^n$ .

EXEMPLO 3.8. Em teoria da aproximação, um problema comum é a aproximação de funções por certas classes de funções especiais, como em séries de Fourier ou séries de polinômios. Séries de Fourier podem ser vistas como bases ortonormais em relação ao produto interno de  $L^2$ . Há, também, vários polinômios ortogonais, dependendo do domínio espacial considerado e produto interno utilizado, que pode ser o próprio produto interno de  $L^2$  ou em produtos internos com peso. Por exemplo, os *polinômios de Legendre* são polinômios considerados no intervalo  $-1 \leq x \leq 1$ , ortogonais em relação ao produto interno de  $L^2(-1, 1)$ . Os *polinômios de Chebyshev* são polinômios considerados também em  $-1 \leq x \leq 1$ , mas são ortogonais em  $L^2_\varphi(-1, 1)$ , com peso  $\varphi(x) = (1 - x^2)^{-1/2}$ . Já os *polinômios de Hermite* são polinômios considerados na reta  $\mathbb{R}$ , ortogonais em  $L^2_\varphi(\mathbb{R})$ , com peso  $\varphi(x) = e^{-x^2}$ . Entre vários outros.

EXEMPLO 3.9. Considere a base de  $\mathbb{R}^3$  composta por

$$\mathcal{a} = \{(1, 1, 0), (0, 1, 1), (1, 0, 1)\}.$$

Aplicando o processo de ortogonalização de Gram-Schmidt, obtemos primeiro o vetor

$$\mathbf{w}_1 = \frac{\sqrt{2}}{2}(1, 1, 0),$$

O segundo vetor é obtido primeiro formando o vetor

$$\begin{aligned}\mathbf{w}'_2 &= (0, 1, 1) - \frac{1}{2}(0, 1, 1) \cdot (1, 1, 0)(1, 1, 0) \\ &= (0, 1, 1) - \frac{1}{2}(1, 1, 0) \\ &= \frac{1}{2}(-1, 1, 2).\end{aligned}$$

Normalizando esse vetor, obtemos o segundo vetor da base ortonormal:

$$\mathbf{w}_2 = \frac{\sqrt{6}}{6}(-1, 1, 2).$$

Para o último vetor, primeiro calculamos

$$\begin{aligned}\mathbf{w}'_3 &= (1, 0, 1) - \frac{1}{2}(1, 0, 1) \cdot (1, 1, 0)(1, 1, 0) - \frac{1}{6}(1, 0, 1) \cdot (-1, 1, 2)(-1, 1, 2) \\ &= (1, 0, 1) - \frac{1}{2}(1, 1, 0) - \frac{1}{6}(-1, 1, 2) \\ &= (1 - 1/2 + 1/6, 0 - 1/2 - 1/6, 1 - 0 - 2/6) \\ &= \frac{2}{3}(1, -1, 1).\end{aligned}$$

Normalizando, obtemos

$$\mathbf{w}_3 = \frac{\sqrt{3}}{3}(1, -1, 1).$$

Assim, obtemos a base ortonormal

$$\mathcal{B} = \left\{ \frac{\sqrt{2}}{2}(1, 1, 0), \frac{\sqrt{6}}{6}(-1, 1, 2), \frac{\sqrt{3}}{3}(1, -1, 1) \right\}.$$

#### 4. Realização do dual e da adjunta via produto interno real

**4.1. Realização do dual via produto interno.** Vimos, na [Seção 3.4](#), a ideia de realização do dual, em que identificamos um dual  $X^*$  com um outro espaço  $Y$ , através de um isomorfismo  $J : X^* \rightarrow Y$ , e com uma forma bilinear  $((\cdot, \cdot)) : X \times Y \rightarrow \mathbb{K}$  tal que

$$\langle \mathbf{f}, \mathbf{u} \rangle = ((\mathbf{u}, J(\mathbf{f}))),$$

para todo  $\mathbf{u} \in X$  e  $\mathbf{f} \in X^*$ . Vamos ver agora, que, graças ao produto interno, podemos identificar o dual  $X^*$  de um espaço vetorial  $X$  com produto interno com o próprio espaço  $X$  e com a forma bilinear sendo o próprio produto interno. Isso também nos permitirá representar a adjunta de um operador linear em  $X$  como sendo também

um operador linear em  $X$ , o que nos dará uma grande ferramenta para analisar operadores com alguma estrutura especial relacionada a sua adjunta (e.g. operadores autoadjuntos, anti-adjuntos, normais, unitários, etc.).

Esse importante resultado é conhecido como o *Teorema da Representação de Riesz*. Vamos considerar apenas a sua versão em dimensão finita.

Há várias maneiras de se demonstrar esse teorema. Veremos duas, aqui. Uma usando o conceito de anulador e baseada nas ideias desenvolvidas no [Capítulo 2](#) e outra usando ortogonalidade, baseada nas ideias desenvolvidas no corrente capítulo.

**TEOREMA 4.1** (Teorema da Representação de Riesz em dimensão finita). *Seja  $\mathbf{f} \in X^*$  um funcional linear em um espaço vetorial real  $X$  de dimensão finita, munido de um produto interno  $((\cdot, \cdot))$ . Então existe um único  $\mathbf{v} \in X$  tal que*

$$\langle \mathbf{f}, \mathbf{u} \rangle = ((\mathbf{u}, \mathbf{v})),$$

para todo  $\mathbf{u} \in X$ .

**DEMONSTRAÇÃO DO TEOREMA 4.1 VIA ANULADOR.** Primeiramente, observe que todo  $\mathbf{v} \in X$  gera um funcional linear  $F(\mathbf{v}) \in X^*$  dado por

$$\langle F(\mathbf{v}), \mathbf{u} \rangle = ((\mathbf{u}, \mathbf{v})).$$

Como o produto interno é bilinear, em particular linear em  $\mathbf{u}$ , segue que  $F(\mathbf{v})$  é, de fato, um funcional linear. É fácil ver que  $F : X \rightarrow X^*$  é injetivo, pois

$$\begin{aligned} F(\mathbf{v}) = F(\mathbf{w}) &\Leftrightarrow \langle F(\mathbf{v}), \mathbf{u} \rangle = \langle F(\mathbf{w}), \mathbf{u} \rangle, \forall \mathbf{u} \in X \\ &\Leftrightarrow ((\mathbf{u}, \mathbf{v})) = ((\mathbf{u}, \mathbf{w})), \forall \mathbf{u} \in X \\ &\Leftrightarrow ((\mathbf{u}, \mathbf{v} - \mathbf{w})) = 0, \forall \mathbf{u} \in X. \end{aligned}$$

Como isso deve valer para todo  $\mathbf{u}$ , escolhemos  $\mathbf{u} = \mathbf{v} - \mathbf{w}$  e deduzimos que

$$\|\mathbf{v} - \mathbf{w}\|^2 = ((\mathbf{v} - \mathbf{w}, \mathbf{v} - \mathbf{w})) = ((\mathbf{v} - \mathbf{w}, \mathbf{u})) = 0.$$

Logo, pela positividade do produto interno, segue que  $\mathbf{v} - \mathbf{w} = \mathbf{0}$ , ou seja,  $\mathbf{v} = \mathbf{w}$ , mostrando que  $F$  é injetivo.

A injetividade de  $F$  significa que o vetor  $\mathbf{v}$  do enunciado do teorema é único, caso exista. Falta mostrar que ele existe. Em outras palavras, falta mostrar que  $F : X \rightarrow X^*$  é sobrejetivo.

Considere, então, a imagem de  $F$ . Observe que  $F$  é uma transformação linear, já que o produto interno é bilinear, portanto linear em  $\mathbf{v}$  também. Com isso, a imagem de  $F$  é um subespaço linear de  $X^*$ ,

$$F = \text{Im } F = \{F(\mathbf{v}); \mathbf{v} \in X\} = \{\mathbf{f} \in X^*; \exists \mathbf{v} \in X, \langle \mathbf{f}, \mathbf{u} \rangle = ((\mathbf{u}, \mathbf{v})), \forall \mathbf{u} \in X\}.$$

Considere, agora, o anulador de  $F$ , que é um subespaço do bidual de  $X$ . Como o bidual de um espaço de dimensão finita pode ser identificado com o próprio espaço (veja [Teorema 4.1](#)), através do isomorfismo canônico  $J : X \rightarrow X^*$  dado por

$$\langle J(\mathbf{u}), \mathbf{f} \rangle_{X^{**}, X^*} = \langle \mathbf{f}, \mathbf{u} \rangle_{X^*, X},$$

podemos considerar o anulador  $F^0$  como um subespaço de  $X$ , dado por

$$F^0 = \{\mathbf{u} \in X; \langle \mathbf{f}, \mathbf{u} \rangle_{X^*, X} = 0, \forall \mathbf{f} \in F\}.$$

Pela definição de  $F$  como imagem de  $F$ , podemos escrever esse anulador na forma

$$F^0 = \{\mathbf{u} \in X; ((\mathbf{u}, \mathbf{v})) = 0, \forall \mathbf{v} \in X\}.$$

Agora, se  $\mathbf{u} \in F^0$ , então, escolhendo  $\mathbf{v} = \mathbf{u}$ , obtemos

$$\|\mathbf{u}\|^2 = ((\mathbf{u}, \mathbf{v})) = \langle F(\mathbf{v}), \mathbf{u} \rangle = 0,$$

ou seja,  $\mathbf{u} = \mathbf{0}$ , mostrando que  $F^0 = \{\mathbf{0}\}$ . Portanto  $\dim(F^0) = 0$ .

Em dimensão finita, segue do [Teorema 5.1](#) que

$$\dim(X^*) = \dim(F) + \dim(F^0) = \dim(F).$$

Como  $F$  é um subespaço de  $X^*$  com a mesma dimensão finita de  $X^*$ , então necessariamente  $F = X^*$ , completando a demonstração de que  $F$  é sobrejetivo e de que sempre existe um único  $\mathbf{v}$  representando um dado funcional linear  $\mathbf{f}$ .  $\square$

**DEMONSTRAÇÃO DO [TEOREMA 4.1](#) VIA ORTOGONALIDADE.** Seja  $\mathbf{f} \in X^*$ . Se  $\mathbf{f} = \mathbf{o}$ , então basta considerar  $\mathbf{v} = \mathbf{0}$  que temos

$$\langle \mathbf{f}, \mathbf{u} \rangle = \langle \mathbf{o}, \mathbf{u} \rangle = 0 = ((\mathbf{u}, \mathbf{0})) = ((\mathbf{u}, \mathbf{v})),$$

para todo  $\mathbf{u} \in X$ . Suponha, então,  $\mathbf{f} \neq \mathbf{o}$ . Considere o núcleo de  $\mathbf{f}$ ,

$$S = \ker \mathbf{f} = \{\mathbf{u} \in X; \langle \mathbf{f}, \mathbf{u} \rangle = 0\},$$

que é um hiperplano em  $X$ , i.e. um subespaço vetorial de co-dimensão um. Como  $X$  é de dimensão finita, então  $S$  também é de dimensão finita (veja [Seção 2](#)). Graças ao [Teorema 3.4](#), podemos considerar a projeção ortogonal  $P_S$  de  $X$  sobre  $S$  e temos, também, que

$$X = S \oplus S^\perp.$$

Como  $S$  tem co-dimensão um, segue que  $S^\perp$  tem dimensão um. Seja  $\mathbf{w} \in S^\perp$  com norma unitária,

$$\|\mathbf{w}\| = 1.$$

Queremos usar  $\mathbf{w}$  para representar  $\mathbf{f}$ . De fato, é natural esperar que  $\mathbf{f}$  seja identificado com um múltiplo de  $\mathbf{w}$ , já que  $S^\perp$  é um complementar do núcleo de  $\mathbf{f}$  e é unidimensional. Não há saída.  $\mathbf{f}$  não se anula em  $S^\perp$  e deve ser uma função linear

unidimensional nesse subespaço, ou seja, se  $x\mathbf{w} \in S^\perp = \text{span}\{\mathbf{w}\}$ , então devemos ter  $x \mapsto \langle \mathbf{f}, x\mathbf{w} \rangle = x\langle \mathbf{f}, \mathbf{w} \rangle = \alpha x$  linear em  $x$ , com algum coeficiente  $\alpha$ . Só devemos encontrar esse coeficiente linear. Para isso, observe que, se  $\mathbf{f}$  está representado por  $\alpha\mathbf{w}$  para algum  $\alpha$ , então

$$\langle \mathbf{f}, \mathbf{u} \rangle = ((\mathbf{u}, \alpha\mathbf{w})),$$

para todo  $\mathbf{u} \in X$ . Em particular,

$$\langle \mathbf{f}, \mathbf{w} \rangle = ((\mathbf{w}, \alpha\mathbf{w})) = \alpha\|\mathbf{w}\|^2 = \alpha.$$

Portanto, devemos ter

$$\mathbf{v} = \langle \mathbf{f}, \mathbf{w} \rangle \mathbf{w}.$$

Vamos mostrar que, de fato,  $\mathbf{f}$  é representado por esse  $\mathbf{v}$ , via produto interno.

Para todo  $\mathbf{u} \in X$ , podemos escrever, de acordo com a decomposição  $X = S \oplus S^\perp$ ,

$$\mathbf{u} = P_S\mathbf{u} + (I - P_S)\mathbf{u}.$$

O operador  $(I - P_S)$  é, precisamente, a projeção ortogonal sobre o perpendicular,  $P_{S^\perp}$ , de modo que podemos escrever

$$\mathbf{u} = P_S\mathbf{u} + P_{S^\perp}\mathbf{u}.$$

Como  $S^\perp$  é unidimensional e  $\mathbf{w}$  é um vetor unitário em  $S^\perp$ , podemos escrever essa projeção ortogonal explicitamente como

$$(I - P_S)\mathbf{u} = P_{S^\perp}\mathbf{u} = ((\mathbf{u}, \mathbf{w}))\mathbf{w}.$$

Com isso,

$$\langle \mathbf{f}, \mathbf{u} \rangle = \langle \mathbf{f}, P_S\mathbf{u} \rangle + \langle \mathbf{f}, P_{S^\perp}\mathbf{u} \rangle = \langle \mathbf{f}, P_S\mathbf{u} \rangle + \langle \mathbf{f}, ((\mathbf{u}, \mathbf{w}))\mathbf{w} \rangle.$$

Como  $P_S\mathbf{u} \in S$ , que é o núcleo de  $\mathbf{f}$ , o primeiro termo se anula e ficamos com

$$\langle \mathbf{f}, \mathbf{u} \rangle = \langle \mathbf{f}, ((\mathbf{u}, \mathbf{w}))\mathbf{w} \rangle.$$

Usando a linearidade do produto de dualidade e do produto interno, em relação o segundo argumento, podemos escrever

$$\langle \mathbf{f}, \mathbf{u} \rangle = ((\mathbf{u}, \mathbf{w}))\langle \mathbf{f}, \mathbf{w} \rangle = ((\mathbf{u}, \langle \mathbf{f}, \mathbf{w} \rangle \mathbf{w})) = ((\mathbf{u}, \mathbf{v})).$$

Como isso vale para todo  $\mathbf{u} \in X$ , isso mostra que  $\mathbf{f}$  é representado por  $\mathbf{v}$ , através do produto interno. A unicidade segue da mesma forma que na demonstração anterior, i.e. se

$$\langle \mathbf{f}, \mathbf{u} \rangle = ((\mathbf{u}, \mathbf{v})) = ((\mathbf{u}, \mathbf{v}')),$$

para todo  $\mathbf{u} \in X$ , para certos  $\mathbf{v}, \mathbf{v}' \in X$ , então

$$((\mathbf{u}, \mathbf{v} - \mathbf{v}')) = 0,$$

para todo  $\mathbf{u} \in X$ , em particular para  $\mathbf{u} = \mathbf{v} - \mathbf{v}'$  e deduzimos que  $\|\mathbf{v} - \mathbf{v}'\|^2 = 0$ , ou seja,  $\mathbf{v}' = \mathbf{v}$ , completando a demonstração.  $\square$

Como consequência do **Teorema 4.1** de Representação Riesz, uma realização do dual de um espaço com produto interno é obtida com o próprio espaço, através do produto interno como produto de dualidade.

**COROLÁRIO 4.1** (Realização do dual de um espaço com produto interno). *Seja  $X$  um espaço vetorial real dimensão finita, munido de um produto interno  $((\cdot, \cdot))$ . Então uma realização do dual é obtida identificando o dual com o próprio espaço e com o produto interno como produto de dualidade, ou seja, existe um isomorfismo linear  $J : X^* \rightarrow X$  tal que*

$$\langle f, \mathbf{u} \rangle = ((\mathbf{u}, J(f))), \quad \forall \mathbf{u} \in X.$$

**DEMONSTRAÇÃO.** Basta tomar  $J = F^{-1}$ , onde  $F : X \rightarrow X^*$  é o isomorfismo linear considerado na demonstração do **Teorema 4.1**.  $\square$

**EXEMPLO 4.1.** Em  $X = \mathbb{R}^2$ , munido do produto escalar, o funcional linear  $f(\mathbf{u}) = f(x, y) = 2x + 3y$ , para  $\mathbf{u} = (x, y)$ , é representado pelo vetor  $\mathbf{v} = (2, 3)$ , visto que

$$\mathbf{v} \cdot \mathbf{u} = (2, 3) \cdot (x, y) = 2x + 3y = f(\mathbf{u}).$$

**EXEMPLO 4.2.** Se consideramos, por outro lado, o produto interno

$$((\mathbf{u}_1, \mathbf{u}_2)) = 2x_1y_1 + 3x_2y_2,$$

para  $\mathbf{u}_1 = (x_1, y_1), \mathbf{u}_2 = (x_2, y_2) \in \mathbb{R}^2$ , então o funcional linear

$$f(\mathbf{u}) = 2x + 3y, \quad \mathbf{u} = (x, y),$$

é representado pelo vetor  $\mathbf{v} = (1, 1)$ , visto que

$$((\mathbf{u}, \mathbf{v})) = 2 \times x \times 1 + 3 \times y \times 1 = 2x + 3y = f(\mathbf{u}).$$

**4.2. Realização da transformação adjunta via produto interno.** A partir do momento em que podemos representar funcionais lineares por elementos do próprio espaço, podemos representar a adjunta  $T^* : Y^* \rightarrow X^*$  de transformações lineares  $T : X \rightarrow Y$  por transformações lineares de  $Y$  em  $X$ . Em alguns textos, apenas essa forma, via produto interno, é considerada, principalmente em espaços de dimensão finita ou quando apenas espaços de Hilbert são considerados. Em outras situações, no entanto, a forma adjunta nos duais é necessária. De qualquer forma, vejamos, aqui, o operador adjunto via produto interno.

**TEOREMA 4.2** (Transformação adjunta em espaços com produto interno). *Sejam  $X$  e  $Y$  espaços vetoriais reais com produto interno e sejam  $((\cdot, \cdot))_X$  e  $((\cdot, \cdot))_Y$  os respectivos produtos internos. Seja  $T \in \mathcal{L}(X, Y)$  uma transformação linear de  $X$  em*

$Y$ . Então existe uma transformação linear  $\mathsf{T}^* \in \mathcal{L}(Y, X)$ , dita a **adjunta** de  $\mathsf{T}$  (via produto interno) que satisfaz

$$((\mathsf{T}\mathbf{u}, \mathbf{v}))_Y = ((\mathbf{u}, \tilde{\mathsf{T}}^*\mathbf{v}))_X,$$

para todo  $\mathbf{u} \in X$ ,  $\mathbf{v} \in Y$ .

OBSERVAÇÃO 4.1. Considere os isomorfismos  $J_X : X^* \rightarrow X$  e  $J_Y : Y^* \rightarrow Y$  obtidos da realização dos duais  $X^*$  e  $Y^*$ , respectivamente, pelos espaços  $X$  e  $Y$  através dos seus produtos internos. Vejamos como a transformação  $\mathsf{T}^* : Y \rightarrow X$  definida no Teorema 4.2 representa a transformação adjunta dada na Definição 4.1, através desses isomorfismos. Para efeito de clareza, vamos denotar, temporariamente, a adjunta de  $Y^*$  em  $X^*$  por  $\tilde{\mathsf{T}}^* : Y^* \rightarrow X^*$ . Temos, por um lado, que

$$\langle \tilde{\mathsf{T}}^*\mathbf{g}, \mathbf{u} \rangle_{X^*, X} = ((\mathbf{u}, J_X \tilde{\mathsf{T}}^*\mathbf{g})).$$

Por outro lado,

$$\langle \tilde{\mathsf{T}}^*\mathbf{g}, \mathbf{u} \rangle_{X^*, X} = \langle \mathbf{g}, \mathsf{T}\mathbf{u} \rangle_{Y^*, Y} = ((\mathsf{T}\mathbf{u}, J_Y(\mathbf{g})))_Y = ((\mathbf{u}, \mathsf{T}^*J_Y(\mathbf{g})))_X.$$

Como isso vale para quaisquer  $\mathbf{g} \in Y^*$  e  $\mathbf{u} \in X$ , temos  $J_X \tilde{\mathsf{T}}^* = \mathsf{T}^*J_Y$ , ou  $\mathsf{T}^* = J_X \tilde{\mathsf{T}}^* J_Y^{-1}$ . Nesse sentido,  $\mathsf{T}^*$  é uma representação de  $\tilde{\mathsf{T}}^*$ , ligada às representações  $J_X$  e  $J_Y$  dos duais de  $X$  e  $Y$ .

OBSERVAÇÃO 4.2. Por conta da Observação 4.1 e para simplificar a notação, utilizamos  $\mathsf{T}^*$  tanto para a adjunta de  $Y^*$  em  $X^*$  quanto para a transformação de  $Y$  em  $X$  que representa a adjunta e chamamos ambas de adjunta de  $\mathsf{T}$ , a menos que seja necessário fazer distinção.

DEFINIÇÃO 4.1 (Transformação adjunta em espaços com produto interno). Sejam  $X$  e  $Y$  espaços vetoriais reais com produto interno e sejam  $((\cdot, \cdot))_X$  e  $((\cdot, \cdot))_Y$  os respectivos produtos internos. Seja  $\mathsf{T} \in \mathcal{L}(X, Y)$  uma transformação linear de  $X$  em  $Y$ . Então chamamos, também, de **transformação adjunta** de  $\mathsf{T}$ , e com a mesma notação, a transformação  $\mathsf{T}^* : Y \rightarrow X$  definida por

$$((\mathsf{T}\mathbf{u}, \mathbf{v}))_Y = ((\mathbf{u}, \mathsf{T}^*\mathbf{v}))_X,$$

para todo  $\mathbf{u} \in X$ ,  $\mathbf{v} \in Y$ , e que representa a transformação adjunta de  $Y^*$  em  $X^*$  conforme explicado na Observação 4.1.

EXEMPLO 4.3. Sejam  $X = \mathbb{R}^3$  e  $Y = \mathbb{R}^2$ , munidos de seus produtos escalares, e seja  $\mathsf{T}(x, y, z) = (x + y, y + z)$  a transformação considerada no Exemplo 4.2. Observe

que

$$\begin{aligned} \mathbb{T}(x, y, z) \cdot (u, v) &= (x + y, y + z) \cdot (u, v) = (x + y)u + (y + z)v \\ &= xu + y(u + v) + zv = (x, y, z) \cdot (u, u + v, v). \end{aligned}$$

Portanto, a adjunta  $\mathbb{T}^* : Y \rightarrow X$  é dada por

$$\mathbb{T}^*(u, v) = (u, u + v, v),$$

de modo que podemos escrever

$$\mathbb{T}(x, y, z) \cdot (u, v) = (x, y, z) \cdot (u, u + v, v) = (x, y, z) \cdot \mathbb{T}^*(u, v).$$

**TEOREMA 4.3** (Representação matricial da adjunta). *Sejam  $X$  e  $Y$  espaços vectoriais reais com produto interno e sejam  $((\cdot, \cdot))_X$  e  $((\cdot, \cdot))_Y$  os respectivos produtos internos. Seja  $\mathbb{T} \in \mathcal{L}(X, Y)$  uma transformação linear de  $X$  em  $Y$ . Considere bases  $\mathfrak{a}$  e  $\mathfrak{b}$  de  $X$  e  $Y$ . Então a matriz  $[\mathbb{T}^*]_{\mathfrak{a}}^{\mathfrak{b}}$  que representa a adjunta  $\mathbb{T}^* : Y \rightarrow X$  entre essas bases é a matriz transposta  $[\mathbb{T}]_{\mathfrak{b}}^{\mathfrak{a}, tr}$  da representação  $[\mathbb{T}]_{\mathfrak{b}}^{\mathfrak{a}}$  de  $\mathbb{T}$  entre essas bases, i.e.*

$$[\mathbb{T}^*]_{\mathfrak{a}}^{\mathfrak{b}} = [\mathbb{T}]_{\mathfrak{b}}^{\mathfrak{a}, tr}$$

**COROLÁRIO 4.2.** *No caso de uma matriz  $\mathbf{A} = (a_{ij})_{ij} \in \mathbb{R}^{m \times n}$  vista como uma transformação  $\mathbb{T}_{\mathbf{A}} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  entre esses espaços munidos do produto escalar, a sua adjunta  $\mathbb{T}_{\mathbf{A}}^* : \mathbb{R}^m \rightarrow \mathbb{R}^n$  está associada à matriz transposta de  $\mathbf{A} = (a_{ji})_{ij}$ .*

## 5. Aplicações

### 5.1. Revisitando a exponencial de operadores.

**5.2. Número de condicionamento e análise de sensibilidade.** Suponha que estejamos estudando um modelo da forma  $\mathbf{v} = \mathbb{T}\mathbf{u}$ , onde dada uma informação de entrada  $\mathbf{u}$ , queiramos saber a resposta  $\mathbf{v}$ . E suponha que, por algum erro de medição ou por alguma incerteza na informação, a entrada avaliada seja perturbada para  $\mathbf{u} + \delta\mathbf{u}$ . Qual o efeito disso na resposta  $\mathbb{T}(\mathbf{u} + \delta\mathbf{u})$ ? Supondo o problema linear, o erro ou a incerteza na resposta é dada por

$$\delta\mathbf{v} = \mathbb{T}(\mathbf{u} + \delta\mathbf{u}) - \mathbb{T}\mathbf{u} = \mathbb{T}\delta\mathbf{u}.$$

Qual é, então, a sensibilidade dessa transformação, em relação a perturbações do dado de entrada?

Considerando um operador  $\mathbb{T} \in \mathcal{L}(X)$  em um espaço normado  $X$ , a norma da transformação nos dá uma estimativa *absoluta* dessa sensibilidade. Nesse contexto, a variação na resposta pode ser estimada pela norma,

$$\|\delta\mathbf{v}\|_X = \|\mathbb{T}\delta\mathbf{u}\|_X \leq \|\mathbb{T}\|_{\mathcal{L}(X)} \|\delta\mathbf{u}\|_X.$$

Ou seja, a perturbação  $\delta \mathbf{u}$  é “amplificada” por um fator  $\|\mathbf{T}\|_{\mathcal{L}(X)}$  (ou “reduzida,” se for uma contração, i.e. se  $\|\mathbf{T}\|_{\mathcal{L}(X)} \leq 1$ ).

Mas as ordens de grandeza de  $\mathbf{u}$  e  $\mathbf{v}$  podem ser bem diferentes e os erros absolutos podem não expressar completamente o efeito da perturbação. Em muitos casos, é mais significativo quantificar as variações relativas.

Ou seja, é mais significativo saber qual é a variação relativa da resposta,

$$\frac{\|\delta \mathbf{v}\|_X}{\|\mathbf{v}\|_X},$$

em comparação com a variação relativa do dado de entrada,

$$\frac{\|\delta \mathbf{u}\|_X}{\|\mathbf{u}\|_X}.$$

Podemos estimar isso via

$$\frac{\|\delta \mathbf{v}\|_X}{\|\mathbf{v}\|_X} = \frac{\|\mathbf{T}\delta \mathbf{u}\|_X}{\|\mathbf{v}\|_X} \leq \frac{\|\mathbf{T}\|_{\mathcal{L}(X)}\|\delta \mathbf{u}\|_X}{\|\mathbf{v}\|_X} = \frac{\|\mathbf{T}\|_{\mathcal{L}(X)}\|\mathbf{u}\|_X}{\|\mathbf{v}\|_X} \frac{\|\delta \mathbf{u}\|_X}{\|\mathbf{u}\|_X}.$$

Caso  $\mathbf{T}$  seja invertível, podemos escrever  $\mathbf{u} = \mathbf{T}^{-1}\mathbf{v}$  e obter

$$\frac{\|\delta \mathbf{v}\|_X}{\|\mathbf{v}\|_X} \leq \frac{\|\mathbf{T}\|_{\mathcal{L}(X)}\|\mathbf{T}^{-1}\mathbf{v}\|_X}{\|\mathbf{v}\|_X} \frac{\|\delta \mathbf{u}\|_X}{\|\mathbf{u}\|_X} \leq \|\mathbf{T}\|_{\mathcal{L}(X)}\|\mathbf{T}^{-1}\|_{\mathcal{L}(X)} \frac{\|\delta \mathbf{u}\|_X}{\|\mathbf{u}\|_X},$$

ou seja,

$$\frac{\|\delta \mathbf{v}\|_X/\|\mathbf{v}\|_X}{\|\delta \mathbf{u}\|_X/\|\mathbf{u}\|_X} \leq \|\mathbf{T}\|_{\mathcal{L}(X)}\|\mathbf{T}^{-1}\|_{\mathcal{L}(X)}.$$

Assim, essa sensibilidade relativa é dada pelo produto

$$\kappa(\mathbf{T}) = \|\mathbf{T}\|_{\mathcal{L}(X)}\|\mathbf{T}^{-1}\|_{\mathcal{L}(X)}.$$

Essa quantidade é um número muito especial chamada de *número de condicionamento* de  $\mathbf{T} \in \mathcal{L}(X)$ .

Observe que o número de condicionamento é simétrico em relação a inversões, i.e.  $\kappa(\mathbf{T}) = \kappa(\mathbf{T}^{-1})$ . Por isso, a mesma análise se aplica ao analisarmos a sensibilidade do *problema inverso*, ou seja, dado  $\mathbf{v}$ , encontrar  $\mathbf{u}$  satisfazendo

$$\mathbf{T}\mathbf{u} = \mathbf{v}.$$

Dada uma variação  $\mathbf{v} + \delta \mathbf{v}$ , o erro relativo da solução  $\mathbf{u} + \delta \mathbf{u}$  de

$$\mathbf{T}(\mathbf{u} + \delta \mathbf{u}) = \mathbf{v} + \delta \mathbf{v}$$

pode ser obtido via

$$\mathbf{u} + \delta \mathbf{u} = \mathbf{T}^{-1}(\mathbf{v} + \delta \mathbf{v}),$$

de modo que

$$\frac{\|\delta\mathbf{u}\|_X/\|\mathbf{u}\|_X}{\|\delta\mathbf{v}\|_X/\|\mathbf{v}\|_X} \leq \|\mathbb{T}^{-1}\|_{\mathcal{L}(X)}\|\mathbb{T}\|_{\mathcal{L}(X)} = \kappa(\mathbb{T}^{-1}) = \kappa(\mathbb{T}).$$

E caso tenhamos uma perturbação nos parâmetros do problema, ou seja, dado uma perturbação  $\delta\mathbb{T}$  no operador, queremos saber a variação na solução  $\mathbf{u} + \delta\mathbf{u}$  do problema

$$(\mathbb{T} + \delta\mathbb{T})(\mathbf{u} + \delta\mathbf{u}) = \mathbf{v}.$$

Nesse caso, temos

$$\mathbf{u} + \delta\mathbf{u} = (\mathbb{T} + \delta\mathbb{T})^{-1}\mathbf{v},$$

de modo que

$$\delta\mathbf{u} = (\mathbb{T} + \delta\mathbb{T})^{-1}\mathbf{v} - \mathbf{u} = (\mathbb{T} + \delta\mathbb{T})^{-1}\mathbf{v} - \mathbb{T}^{-1}\mathbf{v} = ((\mathbb{T} + \delta\mathbb{T})^{-1} - \mathbb{T}^{-1})\mathbf{v}.$$

Usando que  $\mathbf{v} = \mathbb{T}\mathbf{u}$ , o erro relativo pode ser estimado por

$$\frac{\|\delta\mathbf{u}\|_X}{\|\mathbf{u}\|_X} \leq \|(\mathbb{T} + \delta\mathbb{T})^{-1} - \mathbb{T}^{-1}\|_{\mathcal{L}(X)}\|\mathbb{T}\|_{\mathcal{L}(X)}.$$

Ou seja, precisamos analisar, também, a sensibilidade da operação de inversão,

$$\frac{\|(\mathbb{T} + \delta\mathbb{T})^{-1} - \mathbb{T}^{-1}\|_{\mathcal{L}(X)}}{\|\mathbb{T}^{-1}\|_{\mathcal{L}(X)}}.$$

Para isso, primeiro temos que saber se uma perturbação  $\mathbb{T} + \delta\mathbb{T}$  de uma transformação invertível  $\mathbb{T}$  ainda é invertível. Isso pode ser estimado através das normas dos operadores. Primeiramente, escrevemos

$$(\mathbb{T} + \delta\mathbb{T}) = \mathbb{T}(\mathbb{I} + \mathbb{T}^{-1}\delta\mathbb{T}).$$

Assim, vemos que  $\mathbb{T} + \delta\mathbb{T}$  é invertível se, e somente se  $\mathbb{I} + \mathbb{T}^{-1}\delta\mathbb{T}$  for invertível e, nesse caso,

$$(\mathbb{T} + \delta\mathbb{T})^{-1} = (\mathbb{I} + \mathbb{T}^{-1}\delta\mathbb{T})^{-1}\mathbb{T}^{-1}.$$

Com isso,

$$(\mathbb{T} + \delta\mathbb{T})^{-1} - \mathbb{T}^{-1} = ((\mathbb{I} + \mathbb{T}^{-1}\delta\mathbb{T})^{-1} - \mathbb{I})\mathbb{T}^{-1}.$$

Agora, chamando  $\mathbb{R} = -\mathbb{T}^{-1}\delta\mathbb{T}$ , sabemos de [Teorema 1.20](#) que  $\mathbb{I} - \mathbb{R}$  é invertível quando  $\|\mathbb{R}\|_{\mathcal{L}(X)} < 1$ . No nosso caso, podemos estimar

$$\|\mathbb{R}\|_{\mathcal{L}(X)} = \|\mathbb{T}^{-1}\delta\mathbb{T}\| \leq \|\mathbb{T}^{-1}\|_{\mathcal{L}(X)}\|\delta\mathbb{T}\|_{\mathcal{L}(X)} < 1,$$

ou seja, caso  $(\mathbb{T} + \delta\mathbb{T})$  é invertível caso a perturbação satisfaça

$$\|\delta\mathbb{T}\|_{\mathcal{L}(X)} < \frac{1}{\|\mathbb{T}^{-1}\|_{\mathcal{L}(X)}}.$$

Além disso, pela fórmula da inversa dada no **Teorema 1.20**, sabemos que

$$(I - R)^{-1} - I = \sum_{j=1}^{\infty} R^j,$$

de modo que

$$\|(I - R)^{-1} - I\|_{\mathcal{L}(X)} \leq \sum_{j=1}^{\infty} \|R^j\|_{\mathcal{L}(X)} = \frac{\|R\|_{\mathcal{L}(X)}}{1 - \|R\|_{\mathcal{L}(X)}}.$$

Assim, obtemos a estimativa

$$\begin{aligned} \|(\mathbf{T} + \delta\mathbf{T})^{-1} - \mathbf{T}^{-1}\|_{\mathcal{L}(X)} &\leq \|(I + \mathbf{T}^{-1}\delta\mathbf{T})^{-1} - I\|_{\mathcal{L}(X)} \|\mathbf{T}^{-1}\|_{\mathcal{L}(X)} \\ &= \|(I - R)^{-1} - I\|_{\mathcal{L}(X)} \|\mathbf{T}^{-1}\|_{\mathcal{L}(X)} \\ &\leq \frac{\|R\|_{\mathcal{L}(X)}}{1 - \|R\|_{\mathcal{L}(X)}} \|\mathbf{T}^{-1}\|_{\mathcal{L}(X)} \\ &\leq \frac{\|\mathbf{T}^{-1}\|_{\mathcal{L}(X)} \|\delta\mathbf{T}\|_{\mathcal{L}(X)}}{1 - \|\mathbf{T}^{-1}\|_{\mathcal{L}(X)} \|\delta\mathbf{T}\|_{\mathcal{L}(X)}} \|\mathbf{T}^{-1}\|_{\mathcal{L}(X)}. \end{aligned}$$

Combinando com a estimativa anterior, obtemos

$$\frac{\|\delta\mathbf{u}\|_X}{\|\mathbf{u}\|_X} \leq \frac{\|\mathbf{T}^{-1}\|_{\mathcal{L}(X)} \|\delta\mathbf{T}\|_{\mathcal{L}(X)}}{1 - \|\mathbf{T}^{-1}\|_{\mathcal{L}(X)} \|\delta\mathbf{T}\|_{\mathcal{L}(X)}} \|\mathbf{T}\|_{\mathcal{L}(X)} \|\mathbf{T}^{-1}\|_{\mathcal{L}(X)}.$$

Comparando com o erro relativo do operador, obtemos

$$\frac{\|\delta\mathbf{u}\|_X / \|\mathbf{u}\|_X}{\|\delta\mathbf{T}\|_{\mathcal{L}(X)} / \|\mathbf{T}\|_{\mathcal{L}(X)}} \leq \frac{\kappa(\mathbf{T})^2}{1 - \|\mathbf{T}^{-1}\|_{\mathcal{L}(X)} \|\delta\mathbf{T}\|_{\mathcal{L}(X)}}.$$

Limitando a perturbação a

$$\|\mathbf{T}^{-1}\|_{\mathcal{L}(X)} \|\delta\mathbf{T}\|_{\mathcal{L}(X)} \leq \frac{1}{2},$$

chegamos finalmente a estimativa de sensibilidade

$$\frac{\|\delta\mathbf{u}\|_X / \|\mathbf{u}\|_X}{\|\delta\mathbf{T}\|_{\mathcal{L}(X)} / \|\mathbf{T}\|_{\mathcal{L}(X)}} \leq 2\kappa(\mathbf{T})^2.$$

**5.3. Séries de Fourier e aproximações espectrais.** Um série de Fourier, em um intervalo  $[-L/2, L/2]$ , onde  $L > 0$ , tem a forma

$$s(x) = a_0 + \sum_{n=1}^{\infty} a_n \cos\left(\frac{2\pi nx}{L}\right) + \sum_{n=1}^{\infty} b_n \sin\left(\frac{2\pi nx}{L}\right),$$

onde  $a_0, a_n, b_n \in \mathbb{R}$ ,  $n \in \mathbb{N}$ . Essa série está bem definida quando  $(a_n)_n$  e  $(b_n)_n$  são de quadrado somável, i.e.

$$\sum_{n=0}^{\infty} a_n^2 < \infty, \quad \sum_{n=1}^{\infty} b_n^2 < \infty.$$

Nesse caso, a função  $s(x)$  está em  $L^2(-L/2, L/2)$  e vale a *identidade de Parseval*

$$\int_{-L/2}^{L/2} s(x)^2 dx = \sum_{n=0}^{\infty} a_n^2 + \sum_{n=1}^{\infty} b_n^2.$$

Essa identidade tem relação direta com a “fórmula de Pitágoras” e, mais geralmente, com a identidade (3.1). De fato, as somas parciais da série de Fourier podem ser vistas como projeções ortogonais em subespaços senoidais.

Considerando o espaço  $X = L^2(-L/2, L/2)$ , que é um espaço com produto interno

$$((u, v)) = \int_{-L/2}^{L/2} u(x)v(x) dx$$

e norma associada dada por

$$\|u\|^2 = \int_{-L/2}^{L/2} u(x)^2 dx,$$

temos cada  $\cos(2\pi nx/L)$  e  $\sin(2\pi nx/L)$  em  $X$  e gerando subespaços

$$S_m = \left\{ \sum_{n=0}^m (a_n \cos(2\pi nx/L) + b_n \sin(2\pi nx/L)), a_n, b_n \in \mathbb{R} \right\},$$

para  $m = 0, 1, \dots$ , onde incluímos  $b_0$  para simplificar a expressão, mas cujo  $\sin(2\pi jx/L) = 0$  não acrescenta nada para  $j = 0$  e pode ser desprezado. As funções  $\cos(2\pi nx/L)$  e  $\sin(2\pi nx/L)$  são ortogonais entre si, como pode ser verificado via integração por partes. Temos, ainda,

$$\|\cos(2\pi nx/L)\|^2 = \|1\|^2 = L, \quad \text{para } n = 0,$$

e

$$\|\cos(2\pi nx/L)\|^2 = \|\sin(2\pi nx/L)\|^2 = \frac{L}{2}, \quad \text{para } n \in \mathbb{N}.$$

Assim,

$$w_0(x) = \sqrt{\frac{1}{L}}, \quad w_{c,n}(x) = \sqrt{\frac{2}{L}} \cos\left(\frac{2\pi nx}{L}\right),$$

$$w_{s,n}(x) = \sqrt{\frac{2}{L}} \operatorname{sen}\left(\frac{2\pi nx}{L}\right), \quad n = 1, \dots, m,$$

formam uma base ortonormal para o subespaço  $2m + 1$  dimensional  $S_m$ . Dada uma função  $u \in X$ , a projeção ortogonal de  $u$  sobre  $S_m$  é a soma parcial

$$(\mathbf{P}_{S_m} u)(x) = a_0 + \sum_{n=1}^m a_n \cos\left(\frac{2\pi nx}{L}\right) + \sum_{n=1}^m b_n \operatorname{sen}\left(\frac{2\pi nx}{L}\right),$$

onde

$$a_0 = ((u, w_0))w_0(x) = \frac{1}{\|1\|^2}((u, 1)) = \frac{1}{L} \int_{-L/2}^{L/2} u(x) \, dx,$$

$$a_n = ((u, w_{c,n}))w_{c,n}(x) = \frac{2}{L} \int_{-L/2}^{L/2} u(x) \cos\left(\frac{2\pi nx}{L}\right) \, dx,$$

$$b_n = ((u, w_{s,n}))w_{s,n}(x) = \frac{2}{L} \int_{-L/2}^{L/2} u(x) \operatorname{sen}\left(\frac{2\pi nx}{L}\right) \, dx.$$

Conforme válido para qualquer projeção ortogonal, temos

$$\|u - \mathbf{P}_{S_m} u\| \leq \|u - v\|, \quad \forall v \in S_m.$$

A série de Fourier pode ser vista como sendo a fórmula obtida pelo limite das projeções ortogonais,

$$u = \lim_{n \rightarrow \infty} \mathbf{P}_{S_m} u = a_0 + \sum_{n=1}^{\infty} a_n \cos\left(\frac{2\pi nx}{L}\right) + \sum_{n=1}^{\infty} b_n \operatorname{sen}\left(\frac{2\pi nx}{L}\right),$$

com os coeficientes dados em termos de  $u = u(x)$ , como acima.

No caso de uma função ímpar, os coeficientes  $a_n$  se anulam e obtemos uma série de senos, ao passo que no caso de uma função par, os coeficientes  $b_n$  se anulam e obtemos uma série de cossenos. Uma outra forma de enxergar isso é considerar o espaço das funções pares e o subespaço das funções ímpares,

$$A = \{u \in L^2(-L/2, L/2); u(x) = u(-x)\},$$

$$B = \{u \in L^2(-L/2, L/2); u(x) = -u(-x)\}.$$

Temos  $X = A \oplus B$ , com a decomposição

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

Uma observação fundamental é a de que  $A$  e  $B$  são invariantes pelas projeções  $P_{S_m}$ ! Assim, se  $u \in A$ , a sua projeção está em  $S_m \cap A$ , nos dando

$$u = \lim_{n \rightarrow \infty} P_{S_m} u = a_0 + \sum_{n=1}^{\infty} a_n \cos\left(\frac{2\pi nx}{L}\right), \quad \forall u \in A,$$

enquanto que

$$u = \lim_{n \rightarrow \infty} P_{S_m} u = \sum_{n=1}^{\infty} b_n \operatorname{sen}\left(\frac{2\pi nx}{L}\right), \quad \forall u \in B.$$

Uma aplicação importante de séries de Fourier é no estudo de equações a derivadas parciais, já que os subespaços senoidais são invariantes por derivação. Mais ainda, a base ortonormal de senos e cossenos é formada por autovetores do Laplaciano (da segunda derivada, já que estamos considerando, aqui, apenas o caso de uma dimensão espacial)

$$\begin{aligned} \frac{d^2}{dx^2} \cos\left(\frac{2\pi nx}{L}\right) &= -\frac{4\pi n^2}{L^2} \cos\left(\frac{2\pi nx}{L}\right), \\ \frac{d^2}{dx^2} \operatorname{sen}\left(\frac{2\pi nx}{L}\right) &= -\frac{4\pi n^2}{L^2} \operatorname{sen}\left(\frac{2\pi nx}{L}\right). \end{aligned}$$

Ou seja, cada  $w_{c,n}$  e  $w_{s,n}$  é um autovetor do Laplaciano associado ao autovalor

$$\lambda_n = -\frac{4\pi n^2}{L^2},$$

para  $n \in \mathbb{N}$ , sendo  $w_0$  autovetor com autovalor  $\lambda_0 = 0$ .

Por sua vez, a versão complexa desses subespaços, associados à série de Fourier complexa, possui uma base formada por autovetores do próprio operador derivada, de modo que são apropriados para lidar com diversas equações diferenciais parciais. A restrição maior aparece por conta das condições de contorno.

Mas, apenas para efeito de ilustração, considerando a equação a derivadas parciais

$$\frac{\partial u(t, x)}{\partial t} = \nu \frac{\partial^2 u(t, x)}{\partial x^2},$$

com condições de contorno periódicas

$$u(t, x + L) = u(t, x), \quad \forall t \geq 0, x \in \mathbb{R},$$

substituindo  $u(t, \cdot)$  por sua série de Fourier

$$u(t, x) = a_0(t) + \sum_{n=1}^{\infty} a_n(t) \cos\left(\frac{2\pi nx}{L}\right) + \sum_{n=1}^{\infty} b_n(t) \operatorname{sen}\left(\frac{2\pi nx}{L}\right),$$

temos coeficientes dependentes do tempo. Substituindo essa série na expressão para a equação a derivadas parciais e prosseguindo formalmente trocando a ordem do somatório com as derivadas, obtemos um sistema enumerável de equações diferenciais ordinárias para os coeficientes, em termos dos autovalores de cada autovetor,

$$\frac{da_0}{dt} = 0, \quad \frac{da_n}{dt} = -\frac{4\pi n^2}{L^2} a_n, \quad \frac{db_n}{dt} = -\frac{4\pi n^2}{L^2} b_n,$$

para  $n = 1, \dots$ . Assim, podemos, formalmente, escrever a solução da equação a derivadas parciais como

$$u(t, x) = a_0 + \sum_{n=1}^{\infty} a_n e^{-4\pi n^2 t/L^2} \cos\left(\frac{2\pi nx}{L}\right) + \sum_{n=1}^{\infty} b_n e^{-4\pi n^2 t/L^2} \operatorname{sen}\left(\frac{2\pi nx}{L}\right),$$

onde os coeficientes restantes  $a_0 = a(0)$ ,  $a_n = a_n(0)$  e  $b_n = b_n(0)$  são as condições iniciais dos coeficientes dependentes do tempo na expansão de  $u(t, \cdot)$ , ou seja, são os coeficientes da série de Fourier da condição inicial  $u(0, \cdot) = u_0(\cdot)$ .

Um aproximação numérica da solução da equação é obtida restringindo o problema ao subespaço  $S_m$ , para algum  $m \in \mathbb{N}$  relativamente grande. Esse é um método da classe conhecida como *métodos espectrais*.

Métodos espectrais são muito eficientes quando as soluções a serem aproximadas são suaves. Para funções suaves, a taxa de convergência de  $P_m u \rightarrow u$  é bastante rápida, sendo tanto mais rápida quando maior for a diferenciabilidade da função. Já para funções descontínuas, temos o *fenômeno de Gibbs*, com altas oscilações da aproximação próximo ao ponto de descontinuidade e uma lenta taxa de convergência (na norma  $L^2$ , por exemplo, com convergência pontual apenas nos pontos de continuidade da função).

Para exemplificar a lenta convergência, em  $L^2$ , da série de Fourier, considere a função descontínua

$$u(x) = \begin{cases} 1, & x > 0, \\ 0, & x = 0, \\ -1, & x < 0. \end{cases}$$

que é essencialmente a função *senal*  $\operatorname{sgn}(u) = u/|u|$ . Temos, para os coeficientes da série de Fourier no intervalo  $[-1, 1]$  (ou seja,  $L = 2$ ), que  $a_n = 0$ , para  $n = 0, 1, \dots$ ,

visto que a função é ímpar, e com  $b_n$  sendo dado por

$$b_n = \int_0^1 \operatorname{sen}(n\pi x) \, dx = \frac{1}{n\pi} \cos(n\pi x) \Big|_{x=0}^1 = \frac{1}{n\pi} ((-1)^n - 1) = \begin{cases} -\frac{2}{n\pi}, & n \text{ ímpar} \\ 0, & n \text{ par.} \end{cases}$$

Ou seja, os coeficientes  $b_n$  decaem lentamente, em módulo, para zero, de ordem  $\mathcal{O}(1/n)$ , necessitando de um número grande frequências para uma boa aproximação na norma  $L^2$ . Além disso, próximo ao ponto de descontinuidade, há uma grande oscilação da função, conforme ilustrado na Figura...

Por outro lado, caso  $u$  seja continuamente diferenciável, podemos usar integração por partes e obter

$$\begin{aligned} b_n &= \frac{2}{L} \int_{-L/2}^{L/2} u(x) \operatorname{sen} \left( \frac{2\pi nx}{L} \right) \, dx \\ &= \frac{1}{n\pi} u(x) \cos \left( \frac{2\pi nx}{L} \right) \Big|_{x=-L/2}^{L/2} - \frac{1}{n\pi} \int_{-L/2}^{L/2} u'(x) \cos \left( \frac{2\pi nx}{L} \right) \, dx \\ &= -\frac{a'_n}{n\pi}, \end{aligned}$$

onde  $a'_n$  são os coeficientes da série de Fourier da derivada  $u'(x)$  (e onde assumimos que  $u$  é periódica, de modo que  $u(-L/2) = u(L/2)$  e o termo de borda da integração por partes se anula). Ou seja, ganha-se uma potência  $n$ , no denominador, acelerando a taxa de convergência. Mais precisamente, se  $a'_n$  e  $b'_n$  decaem linearmente, então  $a_n$  e  $b_n$  decaem quadraticamente. Esse processo pode ser repetido, caso  $u = u(x)$  tenha derivadas de ordem mais alta, obtendo taxas de convergência cada vez mais altas.

**5.4. Polinômios de Chebyshev e aproximações.** Ao olharmos um rio caudaloso, com um grande fluxo de água, podemos observar que a correnteza diminui rapidamente em locais cada vez mais próximos da margem. Essa região é chamada de *camada limite*, termo utilizado para caracterizar uma região onde há uma grande variação de uma determinada quantidade, separando regiões de comportamentos diferentes. Isso acontece tipicamente próximo à fronteira de uma região, onde alguma condição de contorno é imposta, em contraste com o fenômeno que rege o sistema no interior da região. No caso de um rio, o potencial gravitacional força o escoamento das águas ao longo do leito, mas a margem do rio está fixa e a continuidade do movimento implica no fluxo ter que ser reduzido a zero na fronteira e, portanto, ter uma mudança muitas vezes brusca próximo à margem.

Essa mesma camada limite do movimento do fluido afeta a variação de outras quantidades. Por exemplo, isso explica o motivo de sentirmos mais frio quando bate

um vento. Não necessariamente o vento trás uma massa de ar mais fria do que o ar que está em um determinado ambiente. É a camada limite que nos envolve com um porção de ar que já foi aquecida pelo nosso próprio corpo. Isso causa uma camada limite também para a temperatura, não apenas para o fluxo de ar. Como o nosso corpo é, tipicamente, mais quente que o ambiente, o calor irradiado pelo nosso próprio corpo aquece a pequena camada limite de ar que nos envolve, mantendo essa camada um pouco mais quente que ar ambiente. Quanto bate um vento, o ar nesse camada limite é removido, entrando uma massa de ar relativamente mais fria, nos dando uma sensação de diminuição na temperatura.

Mas o que isso tem a ver com álgebra linear, especialmente com o tópico de ortogonalidade e aproximações de funções? Isso se reflete na escolha das melhores funções em relação às quais devemos aproximar um determinado fenômeno.

Se quisermos aproximar uma função periódica razoavelmente bem comportadas, como em processamento de áudio, por exemplo, as funções trigonométricas da série de Fourier são bastante apropriadas. Mas em problemas com camada limite, será necessária considerar uma aproximação “muito alta”, ou seja, com muitas frequências, em espaços senoidais  $S_m$  com  $m \in \mathbb{N}$  relativamente muito grande, para poder capturar adequadamente a abrupta mudança de valor na quantidade a ser aproximada. Isso está diretamente relacionado ao fenômeno de Gibbs, que acontece em pontos de descontinuidade da função. Nesses casos, pode ser mais eficiente usar outras bases ortonormais para representar a função, permitindo uma aproximação melhor em um subespaço de dimensão relativamente mais baixa. Esse é o caso dos polinômios de Chebyshev.

**5.5. O método de mínimos quadrados.** No *problema de mínimos quadrados*, partimos de espaços  $\mathbb{R}^n$  e  $\mathbb{R}^m$ ,  $n, m \in \mathbb{N}$ , munidos da norma euclidiana, ou, mais geralmente, de espaços  $X$  e  $Y$  munidos de produtos internos  $(\cdot, \cdot)_X$  e  $(\cdot, \cdot)_Y$ , respectivamente. Queremos encontrar uma “melhor solução” para o problema

$$\mathbf{A}\mathbf{u} \approx \mathbf{b}, \quad (5.1)$$

no sentido de minimizar o “erro quadrático”, ou seja, buscamos minimizar a norma quadrática do *erro residual*  $\mathbf{A}\mathbf{u} - \mathbf{b}$  em  $Y$ :

$$\|\mathbf{A}\mathbf{u} - \mathbf{b}\|_Y^2,$$

onde a minimização ocorre em relação à variável  $\mathbf{u} \in X$ . Nesse problema,  $\mathbf{b} \in Y$  e  $\mathbf{A} \in \mathcal{L}(X, Y)$  são fornecidos. Podemos, também, interpretar isso como a “melhor aproximação” para o problema  $\mathbf{A}\mathbf{u} = \mathbf{b}$ . O método de mínimos quadrados passa por observar que, em um ponto de mínimo  $\mathbf{u}$ , qualquer perturbação  $\mathbf{u} + s\mathbf{h}$ , com  $s \neq 0$  e

$\mathbf{h} \in X$ , faz com que

$$\|\mathbf{A}\mathbf{u} - \mathbf{b}\|_Y^2 \leq \|\mathbf{A}(\mathbf{u} + s\mathbf{h}) - \mathbf{b}\|_Y^2, \quad \forall \mathbf{h} \in X, \forall s \in \mathbb{R}.$$

Expandindo a norma quadrática, obtemos

$$\|\mathbf{A}\mathbf{u} - \mathbf{b}\|_Y^2 \leq \|\mathbf{A}\mathbf{u} - \mathbf{b}\|_Y^2 - 2s((\mathbf{A}\mathbf{u} - \mathbf{b}, \mathbf{A}\mathbf{h}))_Y + s^2\|\mathbf{A}\mathbf{h}\|_Y^2.$$

Simplificando e dividindo por  $s$ , obtemos

$$0 \leq -2((\mathbf{A}\mathbf{u} - \mathbf{b}, \mathbf{A}\mathbf{h}))_Y + s\|\mathbf{A}\mathbf{h}\|_Y^2,$$

quando  $s > 0$ , e

$$0 \geq -2((\mathbf{A}\mathbf{u} - \mathbf{b}, \mathbf{A}\mathbf{h}))_Y + s\|\mathbf{A}\mathbf{h}\|_Y^2,$$

quando  $s < 0$ . Fazendo  $s \rightarrow 0$ , sobra

$$((\mathbf{A}\mathbf{u} - \mathbf{b}, \mathbf{A}\mathbf{h}))_Y = 0.$$

para todo  $\mathbf{h} \in X$ . Usando a transformação adjunta, vale a identidade

$$((\mathbf{A}^*(\mathbf{A}\mathbf{u} - \mathbf{b}), \mathbf{h}))_X = ((\mathbf{A}\mathbf{u} - \mathbf{b}, \mathbf{A}\mathbf{h}))_Y = 0.$$

Como isso deve valer para todo  $\mathbf{h} \in X$ , em particular vale para  $\mathbf{h} = \mathbf{A}^*(\mathbf{A}\mathbf{u} - \mathbf{b})$ , o que nos leva a deduzir que o ponto de mínimo  $\mathbf{u}$  satisfaz a equação

$$\mathbf{A}^*(\mathbf{A}\mathbf{u} - \mathbf{b}) = \mathbf{0},$$

em  $X$ . Isso nos leva à *equação normal*

$$\mathbf{A}^*\mathbf{A}\mathbf{u} = \mathbf{A}^*\mathbf{b} \tag{5.2}$$

Em resumo, provamos que se  $\mathbf{u}$  é uma solução da equação normal (5.2), então  $\mathbf{u}$  é um ponto de mínimo do problema de mínimos quadrados (5.1). O *método de mínimos quadrados* consiste, então, em procurar a “melhor aproximação” para (5.1), no sentido dos mínimos quadrados, através da resolução da equação normal (5.2).

## 6. Exercícios

### Exercícios

- 6.1. Para  $1 \leq p, q < \infty$ , mostre que  $\|\cdot\|_{p,q}$  definido por (1.3) é uma norma em  $\mathbb{K}^{m \times n}$ ,  $\mathbb{K}$  real ou complexo.
- 6.2. Seja  $X$  um espaço vetorial real com produto interno  $((\cdot, \cdot))$  e seja  $S \subset X$  um subespaço vetorial de  $X$  de *co-dimensão* finita. Mostre que existe uma projeção ortogonal  $P_S$  sobre  $S$  e que  $X = S \oplus S^\perp$ .

**6.3.** Seja  $X$  um espaço vetorial real com produto interno  $((\cdot, \cdot))$ . Seja  $S \subset X$  um subespaço de  $X$  e suponha que  $Q \in \mathcal{L}(X)$  seja o projetor ortogonal sobre  $S^\perp$ . Mostre que

$$(([\mathbf{u}], [\mathbf{v}]])_{X/S} = ((\mathbf{u} + S, \mathbf{v} + S))_{X/S} = ((Q\mathbf{u}, Q\mathbf{v})), \quad \mathbf{u}, \mathbf{v} \in X,$$

define um produto interno no espaço quociente  $X/S$ , cujos elementos são da forma  $[\mathbf{u}] = \mathbf{u} + S$ ,  $\mathbf{u} \in X$ .



## CAPÍTULO 8

# Teoria Espectral em Espaços Vetoriais Reais com Produto Interno

Um produto interno traz uma estrutura para o espaço que pode ser explorada de diversas formas. Em particular, ela revela propriedades de simetria intrínsecas a vários modelos da natureza e em tecnologia. Muitas dessas simetrias aparecem como simetrias de um operador linear, relacionando o operador e o seu adjunto. Veremos, nesse capítulo, a teoria espectral de operadores com certas simetrias desse tipo, no caso de espaços vetoriais reais com produto interno. O caso complexo será visto no próximo capítulo.

### 1. Operadores especiais com simetria em relação à adjunta

No caso de um operador  $T \in \mathcal{L}(X)$  em um espaço vetorial real com produto interno, o resultado do [Teorema 4.2](#) nos diz que a adjunta  $T^* \in \mathcal{L}(X)$  de  $T$  também é um operador em  $X$ . Como eles vivem no mesmo espaço, podemos classificar alguns operadores de acordo com certas simetrias em relação às suas adjuntas.

**DEFINIÇÃO 1.1.** *Seja  $X$  um espaço vetorial real com produto interno. Seja  $T \in \mathcal{L}(X)$  um operador em  $X$  e considere a sua adjunta  $T^* \in \mathcal{L}(X)$  relativa a esse produto interno.*

- (1) Dizemos que  $T$  é **autoadjunto** quando  $T^* = T$ .
- (2) Dizemos que  $T$  é **anti-autoadjunto** quando  $T^* = -T$ .
- (3) Dizemos que  $T$  é **normal** quando  $TT^* = T^*T$ .
- (4) Dizemos que  $T$  é **ortogonal** quando  $T^*T = TT^* = I$ , ou seja,  $T^* = T^{-1}$ .

Esses operadores aparecem com frequência em aplicações, resultado, em parte, de simetrias que ocorrem nos diversos fenômenos naturais que modelamos. Essas simetrias acarretam em propriedades muitas vezes bastante especiais, como iremos explorar aqui e no próximo capítulo. Em relação a isso, temos a seguinte definição, seguida de uma observação.

DEFINIÇÃO 1.2. Um operador  $T \in \mathcal{L}(X)$  em um espaço vetorial real  $X$  com produto interno é dito **simétrico** quando

$$((Tu, v)) = ((u, Tv)),$$

para  $u, v \in X$  arbitrários, e é dito **anti-simétrico** quando

$$((Tu, v)) = -((u, Tv)),$$

também para  $u, v \in X$  arbitrários.

OBSERVAÇÃO 1.1. As definições de operador simétrico e operador anti-simétrico, vistas na **Definição 1.2**, são equivalentes a dizer que o operador é autoadjunto ou anti-autoadjunto, respectivamente. Em dimensão infinita, onde é comum em aplicações consideramos operadores lineares definidos em um subespaço denso de  $X$ , essas definições não são equivalentes. Mas, aqui, não precisamos nos preocupar com isso.

EXEMPLO 1.1. Em  $\mathbb{R}^2$ , podemos representar um operador qualquer como uma matriz operando na base canônica, da forma

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Para que  $A$  seja simétrica, devemos ter  $c = b$ . Para que  $A$  seja anti-simétrica, devemos ter  $a = d = 0$  e  $c = -b$ . Para que  $A$  seja normal, devemos ter  $c = b$  ou  $c = -b$  e, nesse caso,  $d = a$ . Em resumo

$$\begin{aligned} A &= \begin{bmatrix} a & b \\ b & d \end{bmatrix} && (\text{simétrico}) \Rightarrow \text{autoadjunto} \\ A &= \begin{bmatrix} 0 & b \\ -b & 0 \end{bmatrix} && (\text{anti-simétrico}) \Rightarrow \text{anti-autoadjunto} \\ A &= \begin{bmatrix} a & b \\ b & d \end{bmatrix} \text{ ou } \begin{bmatrix} a & b \\ -b & a \end{bmatrix} && \Rightarrow \text{normal} \end{aligned}$$

### 1.1. Caracterização de projeção ortogonal como projeção autoadjunta.

Com essa representação da adjunta, podemos caracterizar as projeções ortogonais como sendo as projeções *autoadjuntas*.

TEOREMA 1.1. Seja  $X$  um espaço vetorial real com produto interno e seja  $P \in \mathcal{L}(X)$  uma projeção. Então  $P$  é projeção ortogonal se, e somente se,  $P$  é autoadjunta, i.e.  $P^* = P$ .

DEMONSTRAÇÃO. Suponha que  $P$  seja autoadjunta. Se  $u \in \ker(P)$  e  $v \in \text{Im}(P)$ , então temos  $Pu = \mathbf{0}$  e  $v = Pv$ . Com isso,

$$((v, u)) = ((Pv, u)) = ((v, P^*u)) = ((v, Pu)) = ((v, \mathbf{0})) = 0$$

Logo,  $\mathbf{v} \perp \mathbf{u}$ . Como  $\mathbf{u}$  e  $\mathbf{v}$  são arbitrários nesses subespaços, temos  $\ker(P) \perp \text{Im}(P)$ , o que significa que  $P$  é projeção ortogonal.

Agora suponha que  $P$  seja projeção ortogonal. Sejam  $\mathbf{u}, \mathbf{v} \in X$  quaisquer. Então podemos decompor

$$\mathbf{u} = P\mathbf{u} + Q\mathbf{u}, \quad \mathbf{v} = P\mathbf{v} + Q\mathbf{v},$$

onde  $Q = I - P$ . Lembrando que  $\text{Im } P$  e  $\text{Im } Q = \ker P$  são ortogonais, temos

$$\begin{aligned} ((\mathbf{u}, P^*\mathbf{v})) &= ((P\mathbf{u}, \mathbf{v})) = ((P\mathbf{u}, P\mathbf{v} + Q\mathbf{v})) = ((P\mathbf{u}, P\mathbf{v})) \\ &= ((P\mathbf{u} + Q\mathbf{u}, P\mathbf{v})) = ((\mathbf{u}, P\mathbf{v})). \end{aligned}$$

Como  $\mathbf{u}, \mathbf{v} \in X$  são arbitrários, isso significa que  $P^* = P$ .  $\square$

**1.2. Isometria e operadores ortogonais.** Se  $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  é uma base ortonormal em  $\mathbb{R}^n$ , então a matriz  $\mathbf{M} \in \mathbb{R}^{n \times n}$  cujas colunas são formadas por esses vetores, i.e.

$$\mathbf{M} = \begin{bmatrix} | & | & \cdots & | \\ \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_n \\ | & | & \cdots & | \end{bmatrix},$$

é tal que

$$\mathbf{M}^*\mathbf{M} = (\mathbf{w}_i \cdot \mathbf{w}_j)_{ij} = (\delta_{ij})_{ij} = \mathbf{I},$$

ou seja,  $\mathbf{M}$  é uma matriz ortogonal. Além disso,

$$\|\mathbf{M}\mathbf{u}\|^2 = ((\mathbf{M}\mathbf{u}, \mathbf{M}\mathbf{u})) = ((\mathbf{u}, \mathbf{M}^*\mathbf{M}\mathbf{u})) = ((\mathbf{u}, \mathbf{u})).$$

Ou seja,  $\mathbf{M}$  preserva a norma a norma dos vetores. Dito de outra forma,  $\mathbf{M}$  preserva a distância entre vetores do espaço. Nesse sentido, temos o seguinte conceito.

**DEFINIÇÃO 1.3.** *Seja  $F : X \rightarrow X$  uma função (não necessariamente linear) em um espaço métrico (não necessariamente espaço vetorial), com métrica  $d(\cdot, \cdot)$ . Dizemos que  $F$  é uma **isometria** em  $X$  quando*

$$d(F(\mathbf{u}), F(\mathbf{v})) = d(\mathbf{u}, \mathbf{v}), \quad (1.1)$$

para quaisquer  $\mathbf{u}, \mathbf{v} \in X$ .

No caso particular de um espaço vetorial normado e  $F$  não necessariamente linear, a condição (1.1) é equivalente a

$$\|F(\mathbf{u}) - F(\mathbf{v})\| = \|\mathbf{u} - \mathbf{v}\|, \quad (1.2)$$

para  $\mathbf{u}, \mathbf{v} \in X$ . No caso em que a função é uma transformação linear  $F \in \mathcal{L}(X)$ , em um espaço vetorial normado, essa condição é equivalente a

$$\|F(\mathbf{u})\| = \|\mathbf{u}\|, \quad (1.3)$$

para todo  $\mathbf{u} \in X$ .

Vimos acima que um matriz ortogonal é uma isometria. Vamos estender esse resultado para operadores ortogonais quaisquer e incluir uma recíproca desse resultado. Observe que não assumimos inicialmente que  $F$  é linear. Apenas que o espaço  $X$  é um espaço vetorial com produto interno e que  $F$  preserva a origem.

**TEOREMA 1.2.** *Seja  $X$  um espaço vetorial com produto interno e seja  $F : X \rightarrow X$  uma isometria. Então  $F$  é um operador linear ortogonal.*

**DEMONSTRAÇÃO.** Suponha que  $F$  seja linear e ortogonal. Como é linear, temos  $F(\mathbf{0}) = \mathbf{0}$ . Como é ortogonal, temos  $F^*F = I$ , de modo que

$$\|F\mathbf{u}\|^2 = ((F\mathbf{u}, F\mathbf{u})) = ((\mathbf{u}, F^*F\mathbf{u})) = ((\mathbf{u}, \mathbf{u})) = \|\mathbf{u}\|^2,$$

ou seja  $\|F\mathbf{u}\| = \|\mathbf{u}\|$ , para todo  $\mathbf{u}$ , mostrando que  $F$  é um isometria.

Suponha agora que  $F : X \rightarrow X$  preserve a origem e as distâncias. Primeiramente, vamos mostrar que  $F$  também preserva o produto interno, i.e.

$$((F(\mathbf{u}), F(\mathbf{v}))) = ((\mathbf{u}, \mathbf{v})),$$

para  $\mathbf{u}, \mathbf{v} \in X$ . De fato, usando a identidade (2.1) (com  $-\mathbf{v}$  no lugar de  $\mathbf{v}$ ), temos

$$2((F(\mathbf{u}), F(\mathbf{v}))) = \|F(\mathbf{u})\| + \|F(\mathbf{v})\| - \|F(\mathbf{u}) - F(\mathbf{v})\|.$$

Usando que  $F$  preserva a origem, além das distâncias, temos

$$\|F(\mathbf{u})\| = \|F(\mathbf{u}) - \mathbf{0}\| = \|F(\mathbf{u}) - F(\mathbf{0})\| = \|\mathbf{u} - \mathbf{0}\| = \|\mathbf{u}\|.$$

Idem para  $F(\mathbf{v})$ . E também temos

$$\|F(\mathbf{u}) - F(\mathbf{v})\| = \|\mathbf{u} - \mathbf{v}\|.$$

Com isso,

$$2((F(\mathbf{u}), F(\mathbf{v}))) = \|\mathbf{u}\| + \|\mathbf{v}\| - \|\mathbf{u} - \mathbf{v}\|.$$

Usando novamente a identidade (2.1), chegamos a

$$2((F(\mathbf{u}), F(\mathbf{v}))) = 2((\mathbf{u}, \mathbf{v})),$$

mostrando que  $F$  preserva o produto interno entre vetores quaisquer.  $\square$

**OBSERVAÇÃO 1.2** (Isometria preserva ângulos). Observe que, na demonstração do Teorema 1.2, provamos que uma isometria em um espaço com produto interno satisfaz

$$((F(\mathbf{u}), F(\mathbf{v}))) = ((\mathbf{u}, \mathbf{v})), \quad (1.4)$$

para  $\mathbf{u}, \mathbf{v}$  quaisquer, ou seja, uma isometria também preserva ângulos. Como ela também preserva a norma, isso significa que uma isometria preserva os ângulos entre

vetores. Observe, também, que (1.4) inclui a condição de isometria (1.3). Se assumirmos de início que  $F$  é linear, então (1.4) é, de fato, equivalente à definição a  $F$  ser uma isometria.

**COROLÁRIO 1.1.** *Seja  $X$  um espaço vetorial com produto interno e seja  $F : X \rightarrow X$  uma isometria. Então  $F$  é da forma  $F(\mathbf{u}) = \mathbf{u}_0 + M(\mathbf{u} - \mathbf{u}_0)$ , onde  $M \in \mathcal{L}(X)$  é um isometria.*

**TEOREMA 1.3.** *Seja  $M \in \mathcal{L}(X)$  um operador ortogonal em um espaço vetorial real com produto interno. Então o seu determinante tem módulo unitário, i.e.  $\det(M) = \pm 1$ .*

**DEMONSTRAÇÃO.** Sabemos que o determinante do operador adjunto é igual ao do operador (**Teorema 3.8**) e que o determinante de uma composição de operadores é o produto dos determinantes (**Teorema 3.7**). Logo, usando ainda que  $M^*M = I$ , pela definição de operador ortogonal, obtemos que

$$\det(M)^2 = \det(M^*) \det(M) = \det(M^*M) = \det(I) = 1.$$

Portanto,  $\det(M) = \pm 1$ . □

**1.3. Operadores normais e relações de comprimento.** Operadores normais são caracterizados pela seguinte propriedade.

**TEOREMA 1.4.** *Seja  $X$  um espaço vetorial real com produto interno e norma associada  $\|\cdot\|$ . Então  $T \in \mathcal{L}(X)$  é normal se, e somente se,  $\|T^*\mathbf{u}\| = \|T\mathbf{u}\|$ , para todo  $\mathbf{u} \in X$ .*

**DEMONSTRAÇÃO.** Supondo  $T$  normal, temos

$$\begin{aligned} \|T\mathbf{u}\|^2 &= (T\mathbf{u}, T\mathbf{u}) \\ &= (\mathbf{u}, T^*T\mathbf{u}) \\ &= (\mathbf{u}, TT^*\mathbf{u}) \\ &= (T^*\mathbf{u}, T^*\mathbf{u}) \\ &= \|T^*\mathbf{u}\|^2, \end{aligned}$$

provando que  $\|T^*\mathbf{u}\| = \|T\mathbf{u}\|$ , para todo  $\mathbf{u} \in X$ .

Suponha, agora, que essas normas sejam iguais. Então, para  $\mathbf{u}, \mathbf{v} \in X$  arbitrários, usando a identidade (2.1), temos

$$\begin{aligned} ((T^*T\mathbf{u}, \mathbf{v})) &= ((T\mathbf{u}, T\mathbf{v})) \\ &= \frac{1}{2} (\|T\mathbf{u}\|^2 + \|T\mathbf{v}\|^2 - \|T\mathbf{u} - T\mathbf{v}\|^2) \\ &= \frac{1}{2} (\|T^*\mathbf{u}\|^2 + \|T^*\mathbf{v}\|^2 - \|T^*\mathbf{u} - T^*\mathbf{v}\|^2) \\ &= ((T^*\mathbf{u}, T^*\mathbf{v})) \\ &= ((TT^*\mathbf{u}, \mathbf{v})). \end{aligned}$$

Como isso vale para todo  $\mathbf{u}, \mathbf{v} \in X$ , obtemos que  $T^*T = TT^*$ , ou seja, que  $T$  é normal.  $\square$

Também temos a caracterização em relação à comutatividade das partes simétrica e anti-simétrica de  $T$ .

TEOREMA 1.5. *Seja  $X$  um espaço vetorial real com produto interno e seja  $T \in \mathcal{L}(X)$  um operador linear. Sejam*

$$R = \frac{T + T^*}{2}, \quad S = \frac{T - T^*}{2}$$

*as partes simétrica e anti-simétrica de  $T$ , respectivamente. Observe que*

$$T = R + S, \quad R^* = R, \quad S^* = -S.$$

*Então  $T$  é normal se, e somente se,  $R$  e  $S$  comutam entre si.*

DEMONSTRAÇÃO. Observe que

$$(T + T^*)(T - T^*) = T^2 - TT^* + T^*T + T^{*2},$$

ao passo que

$$(T - T^*)(T + T^*) = T^2 + TT^* - T^*T + T^{*2}.$$

Portanto,

$$RS - SR = \frac{1}{4} (2T^*T - 2TT^*).$$

Com isso,

$$RS = SR \Leftrightarrow T^*T = TT^*.$$

Ou seja,  $R$  e  $S$  comutam entre si, se, e somente se,  $T$  é normal.  $\square$

## 2. Análise espectral de operadores autoadjuntos

Um ponto de partida da análise espectral é a existência de autovalores. Nos complexos, isso está garantido pelo **Teorema 3.1**. Esse resultado foi fundamental para a construção da forma de Jordan. Podemos usar a forma de Jordan na análise espectral de um operador qualquer, inclusive de um operador autoadjunto. Mas nem sempre os fins justificando os meios. Vamos ver caminhos mais diretos e mais apropriados para a análise espectral de operadores autoadjuntos. De qualquer forma, a existência de autovalores é fundamental. Como, aqui, estamos considerando operadores em espaços vetoriais reais, precisamos de um resultado de existência de autovalores reais. Para isso, vamos prosseguir de diferentes maneiras. Uma utilizando o resultado do **Teorema 3.1**, via complexificação, e mostrando, na verdade, que o autovalor obtido é real. Outra mostrando diretamente a existência de um autovalor real, via extremos do quociente de Rayleigh-Ritz. A partir disso, o resultado se baseia no fato de que não só os autoespaços são invariantes por um operador autoadjunto, mas também o ortogonal dos autoespaços, permitindo decompor o espaço em subespaços invariantes ortogonais entre si, nos dando uma base ortonormal de autovalores do operador. Esses são os ingredientes principais da diagonalização de um operador:

- (i) Existência de autovalor;
- (ii) Invariância de um subespaço complementar a um autoespaço (que no caso autoadjunto é exatamente o ortogonal do autoespaço);
- (iii) Indução na dimensão, aplicando a hipótese de indução na restrição do operador ao complementar invariante.

**2.1. Existência de autovalores reais.** Primeiro, apresentamos uma demonstração via complexificação, a partir do **Teorema 3.1**.

**TEOREMA 2.1.** *Seja  $X$  um espaço vetorial real com produto interno. Seja  $T \in \mathcal{L}(X)$  um operador autoadjunto em relação a esse produto interno. Então  $T$  possui pelo menos um autovalor real.*

**DEMONSTRAÇÃO DO TEOREMA 2.1 VIA COMPLEXIFICAÇÃO.** Seja  $X_{\mathbb{C}} = X + iX$  a complexificação do espaço real  $X$  e seja  $T_{\mathbb{C}} \in \mathcal{L}(X_{\mathbb{C}})$  a complexificação de  $T$  dada por

$$T_{\mathbb{C}}(\mathbf{u} + i\mathbf{v}) = T\mathbf{u} + iT\mathbf{v},$$

para todo  $\mathbf{u}, \mathbf{v} \in X$ . Como  $X_{\mathbb{C}}$  é complexo, segue do **Teorema 3.1** que  $T_{\mathbb{C}}$  possui pelo menos um autovalor  $\zeta \in \mathbb{C}$ . Vamos mostrar que esse autovalor é real e é um autovalor de  $T$ .

Escreva

$$\zeta = \alpha + i\beta,$$

com  $\alpha, \beta \in \mathbb{R}$ . Seja  $\mathbf{w} = \mathbf{u} + i\mathbf{v}$  um autovetor associado a  $\zeta$ , com  $\mathbf{u}, \mathbf{v} \in X$ . Por um lado,

$$\mathbb{T}_{\mathbb{C}}\mathbf{w} = \mathbb{T}\mathbf{u} + i\mathbb{T}\mathbf{v}.$$

Por outro,

$$\mathbb{T}_{\mathbb{C}}\mathbf{w} = \zeta\mathbf{w} = (\alpha + i\beta)(\mathbf{u} + i\mathbf{v}) = (\alpha\mathbf{u} - \beta\mathbf{v}) + i(\beta\mathbf{u} + \alpha\mathbf{v}).$$

Logo,

$$\mathbb{T}\mathbf{u} = \alpha\mathbf{u} - \beta\mathbf{v}$$

e

$$\mathbb{T}\mathbf{v} = \beta\mathbf{u} + \alpha\mathbf{v}.$$

Muito bem, mas até agora não usamos o fato de que  $\mathbb{T}$  é autoadjunto. Vamos lá. Nesse caso, temos

$$((\mathbb{T}\mathbf{u}, \mathbf{v})) = \alpha((\mathbf{u}, \mathbf{v})) - \beta\|\mathbf{v}\|^2$$

e

$$((\mathbf{u}, \mathbb{T}\mathbf{v})) = \beta\|\mathbf{u}\|^2 + \alpha((\mathbf{v}, \mathbf{u})).$$

Como  $\mathbb{T}$  é autoadjunto, as duas expressões são iguais, ou seja

$$\alpha((\mathbf{u}, \mathbf{v})) - \beta\|\mathbf{v}\|^2 = \beta\|\mathbf{u}\|^2 + \alpha((\mathbf{v}, \mathbf{u})).$$

Os produtos internos se cancelam e sobra

$$\beta(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2) = 0.$$

Como  $\mathbf{w} = \mathbf{u} + i\mathbf{v} \neq \mathbf{0}$ , obtemos que

$$\beta = 0,$$

logo  $\zeta = \alpha$  é real. Junto com isso, obtemos que

$$\mathbb{T}\mathbf{u} = \alpha\mathbf{u}, \quad \mathbb{T}\mathbf{v} = \alpha\mathbf{v}.$$

Como pelo menos um deles é não nulo, obtemos que  $\zeta = \alpha$  é um autovalor real, com uma combinação não nula de  $\mathbf{u}$  e  $\mathbf{v}$  como um autovetor.  $\square$

**OBSERVAÇÃO 2.1.** Uma maneira alternativa de mostrar que o autovalor  $\zeta \in \mathbb{C}$ , do operador complexificado  $\mathbb{T}_{\mathbb{C}}$ , que aparece na demonstração acima do [Teorema 2.1](#), é real, passar por usar que o próprio  $\mathbb{T}_{\mathbb{C}}$  é *hermitiano*, mas isso será feito só na [Seção 4](#).

Vamos, agora, a uma demonstração via extremos do quociente de Rayleigh-Ritz. Essa demonstração usa conceitos de análise real e é baseada na compacidade da bola unitária fechada em espaços de dimensão finita.

A ideia é imaginar uma hipersfera  $S \subset X$ , por exemplo  $S = \{\mathbf{u} \in X; \|\mathbf{u}\| = 1\}$ , sendo deformada em uma elipse  $E = \mathbb{T}(S)$ . O maior semi-eixo é um máximo de

$\|\mathbf{T}\mathbf{u}\|$ , restrito à hipersfera  $S$ . O menor semi-eixo é um mínimo de  $\|\mathbf{T}\mathbf{u}\|$  restrito a  $S$ . Como a função quadrática  $s \mapsto s^2$  é monótona, eles são, respectivamente, pontos de máximo e de mínimo, em  $S$ , de  $\|\mathbf{T}\mathbf{u}\|^2$ , que é diferenciável, conforme comentado na **Observação 2.3**. Assim podemos usar multiplicadores de Lagrange e mostrar que esses pontos críticos são, de fato, autovetores, associados a autovalores reais. O vetor normal a  $\|\mathbf{u}\|^2 = 1$  é  $2\mathbf{u}$ . O vetor normal às curvas de nível de  $((\mathbf{T}\mathbf{u}, \mathbf{u}))$  é  $2\mathbf{T}\mathbf{u}$ . Pela teoria de multiplicadores de Lagrange, um ponto crítico ocorre quando um vetor é um múltiplo do outro, digamos

$$2\mathbf{T}\mathbf{u} = \lambda 2\mathbf{u},$$

com fator multiplicativo  $\lambda$ . Ou seja,

$$\mathbf{T}\mathbf{u} = \lambda\mathbf{u},$$

e  $\lambda$  é um autovalor de  $\mathbf{T}$ .

De outra maneira, como a forma  $((\mathbf{T}\mathbf{u}, \mathbf{u}))$  é quadrática, não precisamos, propriamente, usar diferenciabilidade e o resultado geral de multiplicadores de Lagrange. Podemos mostrar isso mais diretamente, das condições de serem máximo ou mínimo, usando apenas a continuidade. Podemos, também, “liberar”  $\mathbf{u}$  de estar restrito à esfera e considerar o *quociente de Rayleigh-Ritz*

$$\frac{((\mathbf{T}\mathbf{u}, \mathbf{u}))}{\|\mathbf{u}\|^2},$$

definido para  $\mathbf{u} \neq \mathbf{0}$ .

**DEFINIÇÃO 2.1** (Quociente de Rayleigh-Ritz). *Seja  $\mathbf{T} \in \mathcal{L}(X)$  um operador linear em um espaço vetorial com produto interno  $((\cdot, \cdot))$  e norma associada  $\|\cdot\|$ . O **quociente de Rayleigh-Ritz** de  $\mathbf{T}$  é o funcional (não-linear)*

$$\mathfrak{R}(\mathbf{u}) = \frac{((\mathbf{T}\mathbf{u}, \mathbf{u}))}{\|\mathbf{u}\|^2}, \quad (2.1)$$

definido para  $\mathbf{u} \neq \mathbf{0}$ .

Esse quociente é bem comportado, no seguinte sentido.

**TEOREMA 2.2.** *Seja  $\mathbf{T} \in \mathcal{L}(X)$  um operador linear em um espaço vetorial com produto interno e seja  $\mathbf{T} \in \mathcal{L}(X)$ . Suponha que  $X$  seja de dimensão finita (ou, mais geralmente, que  $\mathbf{T}$  seja limitado pela norma induzida pela norma de  $X$ ). Então  $\mathfrak{R} = \mathfrak{R}(\mathbf{u})$ , definido por (2.1), é uma função real contínua e limitada em  $\mathbf{u} \neq \mathbf{0}$ .*

**DEMONSTRAÇÃO.** Sabemos, do **Teorema 1.16**, que  $\mathbf{T}$  é contínua. Pelo **Teorema 1.14**, isso significa dizer que  $\mathbf{T}$  é limitada, no sentido da **Definição 1.10**. Portanto, existe

uma constante  $C \geq 0$  tal que

$$\|\mathbf{T}\mathbf{u}\| \leq C\|\mathbf{u}\|,$$

para todo  $\mathbf{u} \in X$ . Assim, usando a desigualdade de Cauchy-Schwarz [Equation \(2.2\)](#), podemos limitar o quociente de Ritz por

$$|\mathfrak{R}(\mathbf{u})| \leq \frac{|((\mathbf{T}\mathbf{u}, \mathbf{u}))|}{\|\mathbf{u}\|^2} \leq \frac{\|\mathbf{T}\mathbf{u}\|\|\mathbf{u}\|}{\|\mathbf{u}\|^2} \leq C.$$

Agora, sabemos que  $\mathbf{u} \mapsto \mathbf{T}(\mathbf{u})$  é contínuo, que o produto interno (por ser bilinear) é contínuo e que a norma é contínua, de forma que o quociente de Rayleigh-Ritz é contínuo em  $\mathbf{u} \neq \mathbf{0}$ , completando a demonstração.  $\square$

Vamos, agora, a essa outra demonstração do [Teorema 2.1](#).

DEMONSTRAÇÃO DO [TEOREMA 2.1](#) VIA QUOCIENTE DE RAYLEIGH-RITZ. Considere o quociente de Rayleigh-Ritz  $\mathfrak{R}$  definido em [\(2.1\)](#). Pelo [Teorema 2.2](#), esse quociente é contínuo. Pelo [Teorema 1.11](#), o conjunto fechado e limitado  $S = \{\mathbf{u} \in X; \|\mathbf{u}\| = 1\}$  é compacto. Pelo [Teorema 1.12](#), a função  $\mathfrak{R}$  possui um máximo e um mínimo em  $S$ .

Vamos considerar, por exemplo, um ponto de mínimo  $\mathbf{w} \in X$ ,  $\|\mathbf{w}\| = 1$  de  $\mathfrak{R}$  em  $S$  e mostrar que  $\mathbf{w}$  é autovetor de  $\mathbf{T}$ , cujo autovalor é o valor mínimo. Como  $\mathbf{u}/\|\mathbf{u}\|$  pertence a  $S$  para todo  $\mathbf{u} \neq \mathbf{0}$ , podemos escrever

$$\lambda = \mathfrak{R}(\mathbf{w}) = \min_{\mathbf{u} \in S} \mathfrak{R}(\mathbf{u}) = \min_{\mathbf{u} \neq \mathbf{0}} \frac{((\mathbf{T}\mathbf{u}, \mathbf{u}))}{\|\mathbf{u}\|^2}.$$

Seja, então,  $\mathbf{v} \in X$  um vetor não nulo arbitrário e considere  $\mathbf{u} = \mathbf{w} + s\mathbf{v} \in X$  para um escalar  $s$  satisfazendo

$$|s| < \frac{1}{\|\mathbf{v}\|}.$$

Pela continuidade da norma, temos

$$\|\mathbf{u}\| = \|\mathbf{u} - \mathbf{w} + \mathbf{w}\| \geq \|\mathbf{w}\| - \|\mathbf{u} - \mathbf{w}\| = \|\mathbf{w}\| - |s|\|\mathbf{v}\| = 1 - |s|\|\mathbf{v}\| > 0,$$

ou seja,  $\mathbf{u} \neq \mathbf{0}$ . Logo,

$$\mathfrak{R}(\mathbf{w}) = \lambda \leq \frac{((\mathbf{T}\mathbf{u}, \mathbf{u}))}{\|\mathbf{u}\|^2}.$$

Isso nos dá que

$$\lambda\|\mathbf{u}\|^2 \leq ((\mathbf{T}\mathbf{u}, \mathbf{u})).$$

Substituindo  $\mathbf{u}$ , obtemos

$$\lambda(\|\mathbf{w}\|^2 + 2s((\mathbf{w}, \mathbf{v})) + s^2\|\mathbf{v}\|^2) \leq ((\mathbf{T}\mathbf{w}, \mathbf{w})) + 2s((\mathbf{T}\mathbf{w}, \mathbf{v})) + s^2\|\mathbf{v}\|^2.$$

Como  $\|\mathbf{w}\| = 1$  e  $((\mathbb{T}\mathbf{w}, \mathbf{w}))$ , o primeiro termo da expressão à esquerda e o primeiro da expressão à direita se cancelam, sobrando

$$\lambda(2s((\mathbf{w}, \mathbf{v})) + s^2\|\mathbf{v}\|^2) \leq 2s((\mathbb{T}\mathbf{w}, \mathbf{v})) + s^2\|\mathbf{v}\|^2.$$

Considerando  $s > 0$  e dividindo por  $s$ , obtemos

$$2\lambda((\mathbf{w}, \mathbf{v})) + s\lambda\|\mathbf{v}\| \leq 2((\mathbb{T}\mathbf{w}, \mathbf{v})) + s\|\mathbf{v}\|.$$

Fazendo  $s \rightarrow 0$ , obtemos

$$\lambda((\mathbf{w}, \mathbf{v})) \leq 2((\mathbb{T}\mathbf{w}, \mathbf{v})).$$

Considerando  $s < 0$  e dividindo a mesma expressão acima por  $s$ , obtemos a desigualdade reversa

$$2\lambda((\mathbf{w}, \mathbf{v})) + s\lambda\|\mathbf{v}\| \geq 2((\mathbb{T}\mathbf{w}, \mathbf{v})) + s\|\mathbf{v}\|.$$

Fazendo novamente  $s \rightarrow 0$ , agora por números negativos, obtemos

$$\lambda((\mathbf{w}, \mathbf{v})) \geq 2((\mathbb{T}\mathbf{w}, \mathbf{v})).$$

Com as duas desigualdades, obtemos a igualdade

$$\lambda((\mathbf{w}, \mathbf{v})) = ((\mathbb{T}\mathbf{w}, \mathbf{v})),$$

que podemos escrever na forma

$$((\mathbb{T}\mathbf{w} - \lambda\mathbf{w}, \mathbf{v})) = 0.$$

Observe que fizemos isso com  $\mathbf{v}$  arbitrário. Logo, devemos ter

$$\mathbb{T}\mathbf{w} - \lambda\mathbf{w} = \mathbf{0},$$

ou seja,

$$\mathbb{T}\mathbf{w} = \lambda\mathbf{w},$$

com  $\mathbf{w}$  não nulo, mostrando que o valor mínimo  $\lambda$  é autovalor e o ponto de mínimo  $\mathbf{w}$  é um autovetor.  $\square$

**OBSERVAÇÃO 2.2.** A vantagem de usar a diferenciabilidade das funções envolvidas e multiplicadores de Lagrange é mostrar que *qualquer* ponto crítico de  $\mathfrak{R}$  restrito à hipersfera unitária nos dá um autovetor e um autovalor de  $\mathbb{T}$ . De fato, por multiplicadores de Lagrange, os pontos críticos de  $((\mathbb{T}\mathbf{u}, \mathbf{u}))$  restrito a  $S$  são os pontos críticos de

$$\mathfrak{J}(\mathbf{u}, \lambda) = ((\mathbb{T}\mathbf{u}, \mathbf{u})) - \lambda\|\mathbf{u}\|^2.$$

A diferencial de  $\mathbf{u} \mapsto ((\mathbb{T}\mathbf{u}, \mathbf{u}))$  é  $2\mathbb{T}\mathbf{u}$ , visto que  $\mathbb{T}$  é autoadjunto, e a diferencial de  $\mathbf{u} \mapsto \|\mathbf{u}\|^2$  é  $2\mathbf{u}$  (Verifique!). Assim, a diferencial de  $\mathfrak{J} = \mathfrak{J}(\mathbf{u}, \lambda)$  em relação a  $\mathbf{u}$  é

$$D_{\mathbf{u}}\mathfrak{J}(\mathbf{u}, \lambda) = 2\mathbb{T}\mathbf{u} - 2\lambda\mathbf{u},$$

enquanto que

$$D_\lambda \mathfrak{J}(\mathbf{u}, \lambda) = \|\mathbf{u}\|^2.$$

Em um ponto crítico  $(\mathbf{w}, \lambda) \in X \times \mathbb{R}$ , devemos ter

$$D_{\mathbf{u}} \mathfrak{J}(\mathbf{w}) = \mathbf{0}, \quad \|\mathbf{w}\| = 1,$$

ou seja

$$\mathbb{T}\mathbf{w} = \lambda\mathbf{w},$$

e  $\mathbf{w} \neq \mathbf{0}$ , de modo que  $\lambda$  é um autovalor e  $\mathbf{w}$ , um autovetor.

**2.2. Invariância do ortogonal de um autoespaço.** O próximo passo fundamental é a invariância do subespaço ortogonal a um autoespaço do operador autoadjunto, ou, mais geralmente, a um subespaço invariante.

**TEOREMA 2.3.** *Seja  $X$  um espaço vetorial real com produto interno, seja  $\mathbb{T} \in \mathcal{L}(X)$  um operador autoadjunto em relação a esse produto interno e seja  $S \subset X$  um subespaço invariante por  $\mathbb{T}$ . Então o subespaço ortogonal  $S^\perp$  de  $S$  também é invariante por  $\mathbb{T}$ .*

**DEMONSTRAÇÃO.** De fato, se  $\mathbf{v} \in S^\perp$  e  $\mathbf{u} \in S$ , então  $\mathbb{T}\mathbf{u} \in S$ , visto que  $S$  é invariante por  $\mathbb{T}$ , de modo que

$$((\mathbb{T}\mathbf{v}, \mathbf{u})) = ((\mathbf{v}, \mathbb{T}^*\mathbf{u})) = ((\mathbf{v}, \mathbb{T}\mathbf{u})) = 0.$$

Isso mostra que  $\mathbb{T}\mathbf{v} \perp \mathbf{u}$ , para  $\mathbf{u} \in S$  arbitrário, de modo que  $S^\perp$  é invariante por  $\mathbb{T}$ .  $\square$

**2.3. Diagonalização em base ortonormal.** Com a existência de autovalor real garantida e a invariância do subespaço ortogonal a um autoespaço, podemos usar indução e provar a decomposição do espaço em autoespaços ortogonais entre si.

**TEOREMA 2.4.** *Seja  $X$  um espaço vetorial real com produto interno. Seja  $\mathbb{T} \in \mathcal{L}(X)$  um operador autoadjunto em relação a esse produto interno. Então  $X$  possui uma base ortonormal de autovetores de  $\mathbb{T}$ .*

**DEMONSTRAÇÃO.** A demonstração é por indução na dimensão do espaço. Seja  $n = \dim(X)$ . Se  $n = 1$ , então  $X = \text{span}\{\mathbf{w}\}$  para algum  $\mathbf{w} \neq \mathbf{0}$ , que podemos assumir unitário. Todo  $\mathbf{u} \in X$  é um múltiplo de  $\mathbf{w}$ . Além disso,  $\mathbf{w}$  é um autovetor de  $\mathbb{T}$ , com

$$\mathbb{T}\mathbf{w} = ((\mathbb{T}\mathbf{w}, \mathbf{w}))\mathbf{w}.$$

Ou seja,  $X$  possui uma base ortonormal de autovetores de  $\mathbb{T}$ .

Suponha, agora,  $n > 1$ . Graças ao **Teorema 2.1**,  $\mathbb{T}$  possui pelo menos um autovalor real, digamos  $\lambda_1 \in \mathbb{R}$ , com autovetor  $\mathbf{w}_1 \in X$ , que podemos assumir unitário. Como

$S_1 = \text{span}\{\mathbf{w}_1\}$  é invariante, segue de [Teorema 2.3](#) que o subespaço perpendicular  $S_1^\perp$  também é invariante por  $\mathbb{T}$ .

Considere, então, o operador  $\mathbb{T}_1 \mathcal{L}(S_1^\perp)$  dado pela restrição de  $\mathbb{T}$  a  $S_1^\perp$ , que assume valores em  $S_1^\perp$ , já que  $S_1^\perp$  é invariante por  $\mathbb{T}$ . O operador  $\mathbb{T}_1$  também é autoadjunto, agora em  $S_1^\perp$ , munido do produto interno herdado de  $X$ , visto que, para  $\mathbf{u}, \mathbf{v} \in S_1^\perp \subset X$ ,

$$((\mathbb{T}_1 \mathbf{u}, \mathbf{v}))_{S_1^\perp} = ((\mathbb{T} \mathbf{u}, \mathbf{v}))_X = ((\mathbf{u}, \mathbb{T} \mathbf{v}))_X = ((\mathbf{u}, \mathbb{T}_1 \mathbf{v}))_{S_1^\perp}.$$

Como  $\dim(S_1^\perp) = \dim(X) - \dim(S_1) = n - 1$ , a hipótese de indução nos garante que existe uma base ortonormal

$$\{\mathbf{w}_2, \dots, \mathbf{w}_n\}$$

de  $S_1^\perp$  formada por autovetores de  $\mathbb{T}_1$ , que também são autovetores de  $\mathbb{T}$ . Juntando com  $\mathbf{w}_1 \in S_1$ , obtemos uma base ortonormal

$$\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$$

de  $X$ , totalmente formada por autovetores de  $\mathbb{T}$ . Isso completa a demonstração.  $\square$

**OBSERVAÇÃO 2.3.** De uma maneira mais elegante, podemos considerar, ao invés da restrição de  $\mathbb{T}$  a  $S_1^\perp$ , o operador  $\mathbb{T}_1$  induzido por  $\mathbb{T}$  no espaço quociente  $X/S_1$ , que leva um elemento  $[\mathbf{u}] = \mathbf{u} + S_1 \in X/S_1$  no elemento  $[\mathbb{T} \mathbf{u}] = \mathbb{T} \mathbf{u} + S_1 \in X/S_1$ , sendo um operador linear no espaço  $(n - 1)$ -dimensional  $X/S_1$ . Em relação ao produto interno induzido no espaço quociente pelo produto interno de  $X$  (veja [Exercício 6.3](#)), esse operador é autoadjunto, graças ao fato de  $S_1^\perp$  ser invariante também. Veja [Exercício 7.2](#).

Dito de outra forma, temos o seguinte resultado.

**TEOREMA 2.5.** *Seja  $X$  um espaço vetorial real com produto interno. Seja  $\mathbb{T} \in \mathcal{L}(X)$  um operador autoadjunto em relação a esse produto interno. Então  $\mathbb{T}$  possui autovalores reais  $\lambda_1, \dots, \lambda_k$ ,  $k \in \mathbb{N}$ , tais que*

$$X = V(\mathbb{T}, \lambda_1) \oplus \dots \oplus V(\mathbb{T}, \lambda_k)$$

e

$$\mathbb{T} = \lambda_1 \mathbb{P}_1 + \dots + \lambda_k \mathbb{P}_k,$$

onde cada  $\mathbb{P}_j$  é o projetor ortogonal sobre  $V(\mathbb{T}, \lambda_j)$ ,  $j = 1, \dots, k$ .

### 3. Decomposição espectral de operadores anti-autoadjuntos

Como visto na [Definição 1.1](#), operadores anti-autoadjuntos são tais que  $T^* = -T$ . O exemplo clássico de operador anti-autoadjunto nos reais é o produto vetorial em  $\mathbb{R}^3$  com um dado vetor, conforme exemplificado abaixo. Em geral, um operador anti-autoadjunto pode ter um núcleo não trivial e ser composto por uma soma direta de projeções ortogonais bidimensionais com rotações de  $90^\circ$  e homotetias em casa subespaço bidimensional. Uma maneira de obter essa decomposição espectral é associando-o a um operador autoadjunto de maneira apropriada, como veremos aqui.

#### 3.1. Exemplos.

EXEMPLO 3.1 (Produto vetorial como operador linear anti-autoadjunto). Dado um vetor

$$\boldsymbol{\omega} = (a, b, c) \in \mathbb{R}^3,$$

definimos o operador  $T \in \mathcal{L}(\mathbb{R}^3)$  por

$$T\mathbf{u} = \boldsymbol{\omega} \times \mathbf{u} = (bz - cy, cx - az, ay - bx), \quad \mathbf{u} = (x, y, z) \in \mathbb{R}^3.$$

Observe que, para  $\mathbf{u} = (x, y, z)$  e  $\mathbf{v} = (\xi, \eta, \zeta)$ ,

$$\begin{aligned} ((T\mathbf{u}, \mathbf{v})) &= (\boldsymbol{\omega} \times \mathbf{u}) \cdot \mathbf{v} \\ &= (bz - cy)\xi + (cx - az)\eta + (ay - bx)\zeta \\ &= x(c\eta - b\zeta) + y(a\zeta - c\xi) + z(b\xi - a\eta) \\ &= -x(b\zeta - c\eta) - y(c\xi - a\zeta) - z(a\eta - b\xi) \\ &= \mathbf{u} \cdot (\boldsymbol{\omega} \times \mathbf{v}) \\ &= -((\mathbf{u}, T\mathbf{v})). \end{aligned}$$

Ou seja,  $T$  é anti-autoadjunto. Na base canônica  $\mathbf{e}$ , o operador tem a seguinte representação:

$$[T]_{\mathbf{e}} = \begin{bmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{bmatrix}.$$

EXEMPLO 3.2. Em  $\mathbb{R}^2$ , um operador anti-autoadjunto toma a forma (veja [Exemplo 1.1](#))

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

com  $b \in \mathbb{R}$  qualquer. É uma rotação de  $90^\circ$  com uma homotetia com fator  $b$ .

**3.2. Existência de subespaço invariante fundamental.** Uma estratégia para a análise espectral de operadores anti-autoadjuntos é complexificar o espaço  $X$  e o operador  $T$  e observar que  $-iT_{\mathbb{C}}$  é autoadjunto em  $X_{\mathbb{C}}$ , obtendo uma base ortonormal de autovetores complexos, com autovalores reais  $\beta$ , tais que  $-iT_{\mathbb{C}}\mathbf{w} = \beta\mathbf{w}$ . Isso nos dá autovalores imaginários  $i\beta$  para  $T_{\mathbb{C}}$ . Trabalhando com as partes real e imaginária de cada autovalor  $\mathbf{w}$ , obtemos uma representação adequada do operador nos reais, em blocos exatamente da forma vista no **Exemplo 3.2**, i.e.

$$\begin{bmatrix} 0 & -b \\ b & 0 \end{bmatrix}.$$

Também podemos mimetizar essa complexificação trabalhando no produto cartesiano  $X \times X$ , no corpo dos reais. Vamos usar essa ideia para, primeiro, obter um subespaço invariante com esse bloco fundamental. Em seguida, completamos a análise espectral. O caminho via complexos será feito no **Capítulo 9**.

**TEOREMA 3.1.** *Seja  $X$  um espaço vetorial real com produto interno. Seja  $T \in \mathcal{L}(X)$  um operador anti-autoadjunto em relação a esse produto interno. Então existe  $\beta \in \mathbb{R}$  tal que pelo menos uma das alternativas abaixo é válida*

(i)  $\beta = 0$  e existe  $\mathbf{w} \in X$  unitário tal que

$$T\mathbf{w} = \mathbf{0} = 0\mathbf{w},$$

(ii) Ou  $\beta \neq 0$  e existem vetores unitários  $\mathbf{w}_1, \mathbf{w}_2 \in X$  ortogonais entre si tais que

$$T\mathbf{w}_2 = -b\mathbf{w}_1, \quad T\mathbf{w}_1 = b\mathbf{w}_2.$$

**DEMONSTRAÇÃO.** Considere o espaço estendido

$$\tilde{X} = X \times X,$$

e o operador  $\tilde{T} \in \mathcal{L}(\tilde{X})$  definido por

$$\tilde{T}(\tilde{\mathbf{u}}) = (-T\mathbf{u}_2, T\mathbf{u}_1), \quad \tilde{\mathbf{u}} = (\mathbf{u}_1, \mathbf{u}_2) \in \tilde{X} = X \times X.$$

Observe que  $\tilde{T}$  é autoadjunto, pois

$$\begin{aligned} ((\tilde{T}(\tilde{\mathbf{u}}), \tilde{\mathbf{v}}))_{\tilde{X}} &= ((-T\mathbf{u}_2, \mathbf{v}_1))_X + ((T\mathbf{u}_1, \mathbf{v}_2))_X \\ &= ((-\mathbf{u}_2, T^*\mathbf{v}_1))_X + ((\mathbf{u}_1, T^*\mathbf{v}_2))_X \\ &= ((-\mathbf{u}_2, -T\mathbf{v}_1))_X + ((\mathbf{u}_1, -T\mathbf{v}_2))_X \\ &= ((\mathbf{u}_2, T\mathbf{v}_1))_X + ((\mathbf{u}_1, -T\mathbf{v}_2))_X \\ &= ((\mathbf{u}_1, -T\mathbf{v}_2))_X + ((\mathbf{u}_2, T\mathbf{v}_1))_X \\ &= ((\tilde{\mathbf{u}}, \tilde{T}\tilde{\mathbf{v}}))_X. \end{aligned}$$

Com isso, existe pelo menos um autovalor de  $\mathbb{T}$ , que denotamos por  $\beta \in \mathbb{R}$ , com autovetor  $\tilde{\mathbf{w}} = (\mathbf{w}_1, \mathbf{w}_2)$ , de modo que

$$\tilde{\mathbb{T}}(\tilde{\mathbf{w}}) = (-\mathbb{T}\mathbf{w}_2, \mathbb{T}\mathbf{w}_1) = \beta(\mathbf{w}_1, \mathbf{w}_2),$$

ou seja

$$\mathbb{T}\mathbf{w}_2 = -\beta\mathbf{w}_1, \quad \mathbb{T}\mathbf{w}_1 = \beta\mathbf{w}_2.$$

Se  $\beta = 0$ , então fazemos  $\mathbf{w} = \mathbf{w}_1$  e observamos que

$$\mathbb{T}\mathbf{w} = \mathbb{T}\mathbf{w}_1 = \beta\mathbf{w}_2 = \mathbf{0} = 0\mathbf{w},$$

provando o caso 3.1.

Se  $\beta \neq 0$ , então podemos mostrar que  $\mathbf{w}_1$  e  $\mathbf{w}_2$  são ortogonais entre si. De fato, primeiro observe que, para um operador anti-autoadjunto,  $\mathbb{T}\mathbf{u}$  é sempre perpendicular a  $\mathbf{u}$ , visto que

$$((\mathbb{T}\mathbf{u}, \mathbf{u})) = ((\mathbf{u}, \mathbb{T}^*\mathbf{u})) = -((\mathbf{u}, \mathbb{T}\mathbf{u})) = -((\mathbb{T}\mathbf{u}, \mathbf{u})),$$

de modo que

$$((\mathbb{T}\mathbf{u}, \mathbf{u})) = 0.$$

Agora, como estamos assumindo  $\beta \neq 0$ , podemos escrever

$$((\mathbf{u}_1, \mathbf{u}_2)) = \frac{1}{\beta}((\beta\mathbf{u}_1, \mathbf{u}_2)) = \frac{1}{\beta}((\mathbb{T}\mathbf{u}_2, \mathbf{u}_2)) = 0,$$

mostrando que  $\mathbf{u}_1 \perp \mathbf{u}_2$ . Isso completa a demonstração do caso (3.1)  $\square$

**OBSERVAÇÃO 3.1.** Na demonstração do Teorema 3.1, construímos o operador autoadjunto  $\tilde{T}$  no espaço estendido  $\tilde{X} = X \times X$ . Como  $\tilde{T}$  é autoadjunto, existe uma base ortonormal de  $\tilde{X}$  formada por autovetores de  $\tilde{T}$ . Uma outra opção, em princípio mais direta, seria extrair a desejada base ortonormal de  $X$  a partir dessa base ortonormal de  $\tilde{X}$  da decomposição espectral de  $\tilde{T}$ . Porém, como a dimensão de  $\tilde{X}$  é  $n^2$ , onde  $n$  é a dimensão de  $X$ , há muita redundância nessa base, em relação à base necessária para  $X$ . Além do mais, precisamos da combinação de elementos  $\mathbf{w}_1, \mathbf{w}_2$  que estão no espaço produto  $\tilde{X}$ . Ou seja, não é imediato ver como extrair essa base. Deixamos esse caminho para os leitores investigarem. Ao invés de extrairmos diretamente uma base ortonormal de  $X$  a partir de uma base ortonormal de  $\tilde{X}$ , vamos apenas usar essa extensão para extrair um subespaço invariante com o bloco necessário, para então reduzir a dimensão e usar indução para completar a base, conforme feito em diversas outras decomposições.

**3.3. Invariância do ortogonal de um subespaço invariante.** O Teorema 3.1 nos dá um subespaço invariante por  $\mathbb{T}$ , dado por  $S_1 = \text{span}\{\mathbf{w}\}$ , no caso  $\beta = 0$ , e por  $S_1 = \text{span}\{\mathbf{w}_1, \mathbf{w}_2\}$ , no caso  $\beta \neq 0$ . Vamos ver, agora, que o ortogonal desse subespaço também é invariante pelo operador anti-autoadjunto. Na verdade, esse resultado é mais geral, não dependendo da forma do subespaço. Basta que ele seja invariante.

TEOREMA 3.2. *Seja  $X$  um espaço vetorial real com produto interno. Seja  $\mathbb{T} \in \mathcal{L}(X)$  um operador anti-autoadjunto em relação a esse produto interno. Se  $S \subset X$  é um subespaço invariante por  $\mathbb{T}$ , então o subespaço perpendicular  $S^\perp$  também é invariante por  $\mathbb{T}$ .*

DEMONSTRAÇÃO. De fato, se  $\mathbf{v} \in S^\perp$  e  $\mathbf{u} \in S$ , então  $\mathbb{T}\mathbf{u} \in S$ , visto que  $S$  é invariante por  $\mathbb{T}$ , de modo que

$$((\mathbb{T}\mathbf{v}, \mathbf{u})) = ((\mathbf{v}, \mathbb{T}^*\mathbf{u})) = -((\mathbf{v}, \mathbb{T}\mathbf{u})) = 0.$$

Isso mostra que  $\mathbb{T}\mathbf{v} \perp \mathbf{u}$ , para  $\mathbf{u} \in S$  arbitrário, de modo que  $S^\perp$  é invariante por  $\mathbb{T}$ .  $\square$

Aplicando o Teorema 3.2 ao subespaço invariante  $S_1 = \text{span}\{\mathbf{w}\}$ , no caso  $\beta = 0$ , e por  $S_1 = \text{span}\{\mathbf{w}_1, \mathbf{w}_2\}$ , no caso  $\beta \neq 0$ , dados pelo Teorema 3.1, obtemos que  $S_1^\perp$ , em particular, também é invariante por  $\mathbb{T}$ .

**3.4. Decomposição espectral.** De posse dos resultados anteriores, podemos seguir por indução e completar a decomposição espectral de operadores anti-autoadjuntos.

TEOREMA 3.3. *Seja  $X$  um espaço vetorial real com produto interno. Seja  $\mathbb{T} \in \mathcal{L}(X)$  um operador anti-autoadjunto em relação a esse produto interno. Então  $X$  possui uma base ortonormal da forma*

$$\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2,$$

onde

$$\mathcal{B}_1 = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{2k}\}, \quad \mathcal{B}_2 = \{\mathbf{e}_{2k+1}, \dots, \mathbf{w}\},$$

onde  $k$  é um inteiro satisfazendo  $0 \leq 2k \leq n$ , tal que

$$\mathbb{T}\mathbf{w}_{2j} = -\beta_j \mathbf{w}_{2j-1}, \quad \mathbb{T}\mathbf{w}_{2j-1} = \beta_j \mathbf{w}_{2j},$$

para  $j = 1, \dots, k$  e

$$\mathbb{T}\mathbf{w}_j = \mathbf{0},$$



Agora voltemos para  $S_1$ , para completar a base. Se  $\beta = 0$ , completamos a base  $\mathcal{b}'_1$  com  $\mathcal{b}_1$  no final da lista de vetores, obtendo uma base conforme o enunciado.

Se  $\beta \neq 0$ , então  $\mathcal{b}_1 = \{\mathbf{u}_1, \mathbf{u}_2\}$  é uma base ortonormal de  $S_1$ . Completando a base  $\mathcal{b}'_1$  com  $\mathcal{b}_1$  no início da lista de vetores, obtemos uma base ortonormal conforme o enunciado do teorema, completando a demonstração.  $\square$

#### 4. Decomposição espectral de operadores normais

Tendo visto a decomposição espectral de operadores autoadjuntos e anti-autoadjuntos, um caminho natural é observar que um operador normal é a soma de um operador autoadjunto com um anti-autoadjunto que comutam entre si, conforme visto no [Teorema 1.5](#). Graças a essa comutatividade, podemos combinar as decomposições espectrais de cada operador para obter a decomposição espectral de um operador normal.

**4.1. Exemplos.** Em  $\mathbb{R}^2$ , podemos caracterizar um operador normal da seguinte forma.

EXEMPLO 4.1. O exemplo clássico de operador normal é a rotação com homotetia (veja [Exemplo 1.1](#)), dada por

$$\mathbf{A} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}.$$

Pode ser escrita na forma

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

onde

$$M = \sqrt{a^2 + b^2}, \quad \cos \theta = \frac{a}{\sqrt{a^2 + b^2}}, \quad \sin \theta = \frac{b}{\sqrt{a^2 + b^2}}$$

#### 4.2. Decomposição espectral.

TEOREMA 4.1. *Seja  $X$  um espaço vetorial real com produto interno. Seja  $\mathbf{T} \in \mathcal{L}(X)$  um operador normal em relação a esse produto interno. Então  $X$  possui uma base ortonormal da forma*

$$e = \{\{e_1, e_2\}\},$$

onde

$$e_1 = \{\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{2k}\}\}, \quad e_2 = \{\{\mathbf{e}_{2k+1}, \dots, \mathbf{w}\}\},$$

onde  $k$  é um inteiro satisfazendo  $0 \leq 2k \leq n$ , tal que

$$\mathbf{T}\mathbf{w}_{2j} = -\beta_j \mathbf{w}_{2j-1} + \alpha_j \mathbf{w}_{2j}, \quad \mathbf{T}\mathbf{w}_{2j-1} = \alpha_j \mathbf{w}_{2j-1} + \beta_j \mathbf{w}_{2j},$$





### 5. Decomposição em valores singulares (SVD)

Os teoremas espectrais vistos neste capítulo explicam claramente a estrutura de certos operadores com simetrias especiais, como os operadores autoadjuntos, anti-autoadjuntos e, mais geralmente, normais. Os operadores sem essas simetrias, diagonalizáveis ou não, também podem ser bem entendidos via forma canônica de Jordan. Mas como podemos entender melhor transformações lineares entre espaços diferentes? E mesmo no caso geral de operadores, a forma canônica de Jordan não é numericamente estável e também seria importante ter outras decomposições iluminando melhor a estrutura desses operadores. Esse é um dos papéis realizados pela *decomposição em valores singulares (SVD)*.

Uma ideia da SVD passa por observar que um operador linear leva quádricas em quádricas. Em particular, por conta da continuidade, operadores invertíveis levam elipses em elipses (ou elipsóides em elipsóides, em três dimensões, ou os análogos em mais dimensões). Com base nisso, a decomposição em valores singulares se baseia na mudança dos semi-eixos principais da elipse.

Mais explicitamente, isso passa por entender uma transformação linear  $T$  através da estrutura dos conjuntos de nível

$$\|T\mathbf{v}\|^2 = c^2,$$

para  $c \in \mathbb{R}$ . Observe que, pela linearidade, temos

$$\|T(c\mathbf{v})\|^2 = c^2\|T(\mathbf{v})\|^2$$

ou seja, no fundo, basta entender o conjunto de nível

$$\|T\mathbf{v}\|^2 = 1.$$

Considerando o plano  $\mathbb{R}^2$  e usando a norma  $\ell^2$ , esses conjuntos são exatamente quádricas. Em particular, graças à continuidade do operador linear, elipses são levadas em elipses, a menos que o operador seja singular. Nesse sentido, como dito acima, podemos buscar entender a transformação através do que acontece com os eixos principais dessas elipses e do que acontecem com os semi-eixos correspondentes.

No caso de operadores diagonalizáveis, no plano, podemos identificar os eixos principais pelos autoespaços, enquanto que os autovalores indicam a deformação nos semi-eixos. E o que acontece com os operadores não diagonalizáveis e com as transformações mais gerais?

Pensando em uma transformação  $T \in \mathcal{L}(X, Y)$ , observe que

$$\|T\mathbf{v}\|_Y^2 = ((T\mathbf{v}, T\mathbf{v}))_Y = ((T^*T\mathbf{v}, \mathbf{v}))_X.$$

Ou seja, de alguma forma, essa análise passa pelo estudo do operador  $\mathbf{T}^*\mathbf{T} \in \mathcal{L}(X)$ . Esse operador é autoadjunto e não-negativo (veja a demonstração do [Teorema 5.1](#)), o que é explorado como ponto fundamental no entendimento de  $\mathbf{T}$  e, em particular, na decomposição em valores singulares que veremos aqui.

OBSERVAÇÃO 5.1. Mais precisamente, considerando um operador invertível  $\mathbf{T} \in \mathcal{L}(\mathbb{R}^2)$  no plano, escrevendo  $\mathbf{v} = (x, y)$  e  $\mathbf{u} = (\tilde{x}, \tilde{y})$ , a transformação inversa  $\mathbf{v} = \mathbf{T}^{-1}\mathbf{u}$  pode ser escrita na forma

$$\begin{cases} x = a\tilde{x} + b\tilde{y}, \\ y = c\tilde{x} + d\tilde{y}, \end{cases}$$

de modo que uma quádrlica

$$Ax^2 + By^2 + Cz^2 + Dxy + Eyz + Fxz + Gx + Hy + Iz + J = 0,$$

no domínio, é levada em outra quádrlica,

$$\tilde{A}\tilde{x}^2 + \tilde{B}\tilde{y}^2 + \tilde{C}\tilde{z}^2 + \tilde{D}\tilde{x}\tilde{y} + \tilde{E}\tilde{y}\tilde{z} + \tilde{F}\tilde{x}\tilde{z} + \tilde{G}\tilde{x} + \tilde{H}\tilde{y} + \tilde{I}\tilde{z} + \tilde{J} = 0,$$

na imagem. Pela continuidade da transformação, se a quádrlica inicial for uma elipse, a imagem deve ser uma elipse, também (curvas suaves fechadas são levadas em curvas fechadas suaves). Portanto, podemos pensar em entender a transformação através do que acontece com os semi-eixos principais dessas elipses e com os seus semi-eixos. A deformação nos semi-eixos é dada exatamente pelos valores principais.

OBSERVAÇÃO 5.2. A [Observação 5.1](#) se baseia na invertibilidade do operador  $\mathbf{T} \in \mathcal{L}(X, Y)$  para mostrar que quádrlicas em  $X$  (em particular, elipses) são levadas em quádrlicas em  $Y$  (resp. elipses). No caso geral, uma ideia é considerar um complementar  $X'$  ao núcleo de  $\mathbf{T}$  em  $X$  e a imagem  $Y' = \mathbf{T}(X) = \mathbf{T}(X')$  e trabalhar com a restrição de  $\mathbf{T}$  como operador invertível de  $X'$  em  $Y'$ . Em uma forma mais elegante, podemos aplicar a [Observação 5.1](#) ao operador invertível de  $X/\ker(\mathbf{T})$  em  $\text{Im}(\mathbf{T})$  induzido por  $\mathbf{T}$ .

### 5.1. Definição e exemplos.

DEFINIÇÃO 5.1 (Decomposição em Valores Singulares). *Sejam  $X$  e  $Y$  espaços vetoriais reais munidos de produto interno e seja  $\mathbf{T} \in \mathcal{L}(X, Y)$ . Suponha que  $X$  e  $Y$  sejam de dimensão finita, com  $n = \dim(X)$ ,  $m = \dim(Y) \in \mathbb{N}$ . Uma **decomposição em valores singulares** de  $\mathbf{T}$  é composta por bases ortonormais  $\mathfrak{a} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  de  $X$  e  $\mathfrak{b} = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$  de  $Y$  e escalares reais não negativos  $\sigma_1 \geq \dots \geq \sigma_n \geq 0$ , denominados de **valores singulares**, tais que*

$$\mathbf{T}\mathbf{v}_j = \sigma_j\mathbf{u}_j, \quad j = 1, \dots, \min\{n, m\}, \quad (5.1)$$

com

$$\mathbf{v}_j \in \ker(\mathbb{T}), \quad j = m + 1, \dots, n, \quad (5.2)$$

caso  $n > m$ , e

$$\mathbf{u}_j \perp \text{Im}(\mathbb{T}), \quad j = n + 1, \dots, m, \quad (5.3)$$

caso  $m > n$ .

Na versão matricial, a decomposição em valores singulares toma a seguinte forma.

**DEFINIÇÃO 5.2** (Decomposição em Valores Singulares na forma matricial). *Sejam  $n, m \in \mathbb{N}$  e seja  $\mathbf{A} \in \mathbb{R}^{m \times n}$ . Uma **decomposição em valores singulares** de  $\mathbf{A}$  consiste de matrizes ortogonais  $\mathbf{U} \in \mathbb{R}^{n \times n}$  e  $\mathbf{V} \in \mathbb{R}^{m \times m}$  e uma matriz diagonal  $\mathbf{\Sigma} \in \mathbb{R}^{m \times n}$ , com todos os elementos sendo reais não negativos, tais que*

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*.$$

**OBSERVAÇÃO 5.3.** Na **Definição 5.2**, uma matriz  $\mathbf{A}$  gera uma transformação linear  $\mathbb{T}_{\mathbf{A}} \in \mathcal{L}(X, Y)$  entre os espaços reais  $X = \mathbb{R}^n$  e  $Y = \mathbb{R}^m$  munidos do produto escalar canônico. Nesse contexto, uma decomposição em valores singulares de  $\mathbb{T}_{\mathbf{A}}$  consiste de bases ortonormais  $\mathcal{a} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$  de  $\mathbb{R}^n$  e  $\mathcal{b} = \{\mathbf{v}_1, \dots, \mathbf{v}_m\}$  de  $\mathbb{R}^m$ , junto com valores singulares  $\sigma_1 \geq \dots \geq \sigma_n \geq 0$  satisfazendo (5.1).

Considerando a matriz  $\mathbf{V} \in \mathbb{R}^{n \times n}$  cujas colunas são os vetores  $\mathbf{v}_1, \dots, \mathbf{v}_n$ , a matriz  $\mathbf{U} \in \mathbb{R}^{m \times m}$  cujas colunas são os vetores  $\mathbf{u}_1, \dots, \mathbf{u}_m$  e a matriz diagonal  $\mathbf{\Sigma} = (s_{ij})_{ij} \in \mathbb{R}^{m \times n}$  com  $s_{jj} = \sigma_j$ ,  $j = 1, \dots, \min\{n, m\}$  e  $s_{ij} = 0$ ,  $i \neq j$ , a identidade (5.1) significa exatamente que

$$\mathbf{A}\mathbf{V} = \mathbf{U}\mathbf{\Sigma}.$$

Como as bases  $\mathcal{a}$  e  $\mathcal{b}$  são ortonormais, as matrizes  $\mathbf{V}$  e  $\mathbf{U}$  são ortogonais. Com isso,  $\mathbf{V}^{-1} = \mathbf{V}^*$ . Logo, obtemos

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*.$$

Reciprocamente, dada uma decomposição em valores singulares de  $\mathbf{A}$ , encontramos uma decomposição em valores singulares de  $\mathbb{T}_{\mathbf{A}}$ .

**OBSERVAÇÃO 5.4.** Veja, em **Exercício 7.5**, **Exercício 7.6** e **Exercício 7.7**, que, nos casos de operadores autoadjuntos, anti-autoadjuntos e, mais geralmente, normais, que há uma relação direta entre os valores singulares e os autovalores do operador e também entre as bases ortonormais da decomposição SVD e da decomposição espectral.

**EXEMPLO 5.1.** A matriz diagonal

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

tem valores singulares  $\sigma_1 = 3$  e  $\sigma_2 = 2$ . Uma decomposição em valores singulares é obtida escolhendo  $\mathbf{v}_1 = \mathbf{e}_2$ ,  $\mathbf{v}_2 = \mathbf{e}_1$ ,  $\mathbf{u}_1 = -\mathbf{e}_2$  e  $\mathbf{u}_2 = \mathbf{e}_1$ , de modo que

$$\mathbf{A}\mathbf{v}_1 = \mathbf{A}\mathbf{e}_2 = -3\mathbf{e}_2 = 3\mathbf{u}_1, \quad \mathbf{A}\mathbf{v}_2 = \mathbf{A}\mathbf{e}_1 = 2\mathbf{e}_1 = 2\mathbf{u}_2.$$

Em termos matriciais,

$$\mathbf{V} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad \mathbf{\Sigma} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}.$$

EXEMPLO 5.2. A matriz de rotação com homotetia

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

tem autovalores  $\pm 2i$ . Seus valores singulares são  $\sigma_1 = \sigma_2 = 2$ . Uma decomposição em valores singulares é obtida escolhendo  $\mathbf{v}_1 = \mathbf{e}_1$ ,  $\mathbf{v}_2 = \mathbf{e}_2$ ,  $\mathbf{u}_1 = \mathbf{e}_2$  e  $\mathbf{u}_2 = -\mathbf{e}_1$ , de modo que

$$\mathbf{A}\mathbf{v}_1 = \mathbf{A}\mathbf{e}_1 = 2\mathbf{e}_2 = 2\mathbf{u}_1, \quad \mathbf{A}\mathbf{v}_2 = \mathbf{A}\mathbf{e}_2 = -\mathbf{e}_1 = 2\mathbf{u}_2.$$

Em termos matriciais,

$$\mathbf{V} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{\Sigma} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.$$

EXEMPLO 5.3. A matriz

$$\mathbf{A} = \begin{bmatrix} 6 & -2 \\ 7 & 6 \end{bmatrix}$$

tem autovalores complexos

$$\lambda_{\pm} = \frac{-12 \pm \sqrt{56}i}{2} = -6 \pm \sqrt{14}i.$$

Esta não é uma matriz normal. De qualquer maneira, a sua forma canônica real de Jordan é

$$\mathbf{J} = \begin{bmatrix} 6 & -\sqrt{14} \\ \sqrt{14} & 6 \end{bmatrix}.$$

Por outro lado, a sua decomposição em valores singulares é dada pelas matrizes

$$\mathbf{V} = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}, \quad \mathbf{U} = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}, \quad \mathbf{\Sigma} = \begin{bmatrix} 10 & 0 \\ 0 & 5 \end{bmatrix}$$

Observe, ainda, que  $|\lambda_{\pm}| = \sqrt{50} \approx 7.07$ , bem entre os valores singulares  $\sigma_1 = 10$  e  $\sigma_2 = 5$ .

EXEMPLO 5.4. Considere a matriz

$$\mathbf{A} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix}.$$

Observe que

$$\mathbf{A}\mathbf{e}_1^2 = 3\mathbf{e}_1^3, \quad \mathbf{A}\mathbf{e}_2^2 = 2\mathbf{e}_2^3, \quad \mathbf{e}_3^3 \perp \text{Im}(\mathbf{A}).$$

A decomposição em valores singulares de  $\mathbf{A}$  é dada por

$$\alpha = e_2 = \{\mathbf{e}_1^2, \mathbf{e}_2^2\}, \quad \beta = e_3 = \{\mathbf{e}_1^3, \mathbf{e}_2^3, \mathbf{e}_3^3\}, \quad \sigma_1 = 3, \quad \sigma_2 = 2.$$

Em forma matricial, a decomposição em valores singulares de  $\mathbf{A}$  é dada pelas matrizes

$$\mathbf{V} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{\Sigma} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}.$$

EXEMPLO 5.5. Considere a matriz

$$\mathbf{A} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \end{bmatrix}.$$

Observe que

$$\mathbf{A}\mathbf{e}_1^3 = 3\mathbf{e}_1^2, \quad \mathbf{A}\mathbf{e}_2^3 = 2\mathbf{e}_2^2, \quad \mathbf{e}_3^3 \in \ker(\mathbf{A}),$$

onde  $e_n = \{\{\mathbf{e}_1^n, \dots, \mathbf{e}_n^n\}\}$  é a base canônica de  $\mathbb{R}^n$ ,  $n \in \mathbb{N}$ . A decomposição em valores singulares de  $\mathbf{A}$  é dada por

$$\alpha = e_3 = \{\mathbf{e}_1^3, \mathbf{e}_2^3, \mathbf{e}_3^3\}, \quad \beta = e_2 = \{\mathbf{e}_1^2, \mathbf{e}_2^2\}, \quad \sigma_1 = 3, \quad \sigma_2 = 2.$$

Em forma matricial, a decomposição em valores singulares de  $\mathbf{A}$  é dada pelas matrizes

$$\mathbf{V} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{\Sigma} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

OBSERVAÇÃO 5.5. Através da decomposição SVD de uma transformação, podemos obter a decomposição SVD da adjunta e entender melhor a relação entre uma transformação e a sua adjunta. Veja [Exercício 7.8](#).

## 5.2. Demonstração de existência de decomposição SVD.

TEOREMA 5.1 (Decomposição em Valores Singulares). *Sejam  $X$  e  $Y$  espaços vectoriais reais munidos de produto interno e seja  $T \in \mathcal{L}(X, Y)$ . Suponha que  $X$  e  $Y$  sejam de dimensão finita, com  $n = \dim(X), m = \dim(Y) \in \mathbb{N}$ . Então existe uma decomposição em valores singulares de  $T$ , no sentido da [Definição 5.1](#).*

DEMONSTRAÇÃO DO [TEOREMA 5.1](#) VIA OPERADOR DE GRAM. O operador de Gram  $T^*T \in \mathcal{L}(X)$  é autoadjunto e não-negativo, visto que

$$(T^*T)^* = T^*T^{**} = T^*T$$

e

$$((T^*T\mathbf{u}, \mathbf{u})) = \|T\mathbf{u}\|^2 \geq 0.$$

Desse modo,  $T^*T$  possui um base ortonormal de autovetores, que denotamos por

$$\mathcal{a} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}.$$

Como os autovalores são não-negativos, escrevemos os autovalores na forma  $\sigma_j^2$ ,  $j = 1, \dots, n$ , ou seja, definimos  $\sigma_j$  como sendo a raiz quadrada de cada autovalor do operador de Gram  $T^*T$ . Assumimos que os autovetores estão ordenados de acordo com os seus autovalores, do maior para o menor. Ou seja, assumimos que

$$T^*T\mathbf{v}_j = \sigma_j^2\mathbf{v}_j, \quad j = 1, \dots, n,$$

com

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0.$$

Seja  $k \in \mathbb{N}$  o posto de  $T$ , de modo que

$$\sigma_1 \geq \dots \geq \sigma_k > 0 = \sigma_{k+1} = \dots = \sigma_n.$$

Dessa forma, os vetores  $T\mathbf{v}_j$ ,  $j = 1, \dots, k$ , são não nulos, com

$$\|T\mathbf{v}_j\|^2 = ((T\mathbf{v}_j, T\mathbf{v}_j)) = ((T^*T\mathbf{v}_j, \mathbf{v}_j)) = \sigma_j^2((\mathbf{v}_j, \mathbf{v}_j)) = \sigma_j^2\|\mathbf{v}_j\|^2 = \sigma_j^2 > 0,$$

onde usamos que cada  $\mathbf{v}_j$  está normalizado para ter norma unitária. Além disso, os vetores  $\{T\mathbf{v}_1, \dots, T\mathbf{v}_k\}$  formam um conjunto de vetores ortogonais entre si. De fato, temos

$$((T\mathbf{v}_j, T\mathbf{v}_i)) = ((T^*T\mathbf{v}_j, \mathbf{v}_i)) = \sigma_j^2((\mathbf{v}_j, \mathbf{v}_i)) = 0,$$

para  $i \neq j$ , visto que  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  é uma base ortonormal. Definindo

$$\mathbf{u}_j = \frac{1}{\sigma_j} T\mathbf{v}_j, \quad j = 1, \dots, k,$$

temos

$$\|\mathbf{u}_j\|^2 = \frac{1}{\sigma_j^2}((\mathbb{T}\mathbf{v}_j, \mathbb{T}\mathbf{v}_j)) = \frac{1}{\sigma_j^2}((\mathbb{T}^*\mathbb{T}\mathbf{v}_j, \mathbf{v}_j)) = ((\mathbf{v}_j, \mathbf{v}_j)) = \|\mathbf{v}_j\|^2 = 1.$$

Portanto, obtemos uma base ortonormal  $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$  da imagem  $\text{Im } \mathbb{T}$  da transformação  $\mathbb{T}$ . Se necessário, i.e. se  $k < m$ , completamos esse conjunto até uma base ortonormal

$$\mathcal{b} = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$$

de  $Y$ . Dessa forma, obtemos as desejadas bases ortonormais  $\mathcal{a}$  e  $\mathcal{b}$ . Pela própria normalização dos vetores  $\mathbf{u}_j$ , obtemos que

$$\mathbb{T}\mathbf{v}_j = \sigma_j \mathbf{u}_j, \quad j = 1, \dots, k.$$

Caso  $k < n$ , temos, para  $k < j \leq n$ , como  $\sigma_j = 0$ ,

$$\|\mathbb{T}\mathbf{v}_j\|^2 = ((\mathbb{T}\mathbf{v}_j, \mathbb{T}\mathbf{v}_j)) = ((\mathbb{T}^*\mathbb{T}\mathbf{v}_j, \mathbf{v}_j)) = \sigma_j^2((\mathbf{v}_j, \mathbf{v}_j)) = 0,$$

ou seja,  $\mathbb{T}\mathbf{v}_j = \mathbf{0}$  e  $\mathbf{v}_j \in \ker(\mathbb{T})$ .

Caso  $k < m$ , temos, simplesmente,  $\mathbf{u}_j \perp \text{Im}(\mathbb{T})$ .

Note que, caso  $k < \min\{n, m\}$ , então, para  $k < j \leq \min\{n, m\}$ , temos  $\mathbf{v}_j \in \ker(\mathbb{T})$ ,  $\sigma_j = 0$  e  $\mathbf{u}_j \in \text{Im}(\mathbb{T})$  e ainda assim podemos escrever

$$\mathbb{T}\mathbf{v}_j = \mathbf{0} = \sigma_j \mathbf{u}_j,$$

portanto a identidade (5.1) vale para todo  $j = 1, \dots, \min\{n, m\}$ , independentemente do posto de  $\mathbb{T}$ .  $\square$

**OBSERVAÇÃO 5.6.** A demonstração acima do **Teorema 5.1** se baseia na análise espectral do operador de Gram  $\mathbb{T}^*\mathbb{T}$  e obtém os valores singulares como autovalores do operador de Gram e a base  $\mathcal{a} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  como autovetores correspondentes. Em princípio, poderíamos ter decomposições em valores singulares diferentes dessa. No entanto, essa é a única decomposição possível. A única variação que podemos ter é na escolha dos autovetores, dentro de cada autoespaço. Mais precisamente, a única variação possível é na escolha da base ortonormal de cada autoespaço do operador de Gram. Em relação à base  $\mathcal{b} = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ , temos  $\mathbf{u}_j$  determinados pela escolha de  $\mathbf{v}_j$ , nos índices em que  $\sigma_j \neq 0$ , tendo liberdade de escolha apenas na base ortonormal do subespaço perpendicular  $\text{Im}(\mathbb{T})^\perp$ . Veja **Exercício 7.4**.

**OBSERVAÇÃO 5.7.** Uma outra demonstração do **Teorema 5.1** pode ser feita através do operador  $\mathbb{T}\mathbb{T}^* \in \mathcal{L}(Y)$ . Os valores singulares também estão associados aos autovalores desse operador. Veja **Exercício 7.9**.

**5.3. Caracterização minimax.** Vimos, no caso de operadores autoadjuntos, que os autovalores podem ser obtidos via minimização ou maximização dos quocientes de Rayleigh  $((\mathbf{T}\mathbf{u}, \mathbf{u})/\|\mathbf{u}\|^2)$ . No caso da decomposição em valores singulares, trabalhamos com os autovalores do operador de Gram  $\mathbf{T}^*\mathbf{T}$ . Isso nos leva ao quociente de Rayleigh

$$\frac{((\mathbf{T}^*\mathbf{T}\mathbf{u}, \mathbf{u}))_X}{\|\mathbf{u}\|_X^2},$$

que pode ser escrito como

$$\frac{\|\mathbf{T}\mathbf{u}\|_Y^2}{\|\mathbf{u}\|_X^2}.$$

Assim, o maior valor singular pode ser obtido via

$$\sigma_1 = \max_{\mathbf{u} \neq \mathbf{0}} \frac{\|\mathbf{T}\mathbf{u}\|_Y^2}{\|\mathbf{u}\|_X^2},$$

enquanto que o menor valor singular pode ser obtido via

$$\sigma_n = \min_{\mathbf{u} \neq \mathbf{0}} \frac{\|\mathbf{T}\mathbf{u}\|_Y^2}{\|\mathbf{u}\|_X^2}.$$

Valores singulares intermediários podem ser obtidos via fórmulas minimax,

$$\sigma_j = \min_{\dim(S)=j} \max_{\mathbf{u} \in S \setminus \{\mathbf{0}\}} \frac{\|\mathbf{T}\mathbf{u}\|_Y^2}{\|\mathbf{u}\|_X^2},$$

onde  $S$  são subespaços de  $X$ .

A maximização para obter o maior valor singular pode ser vista como um problema de multiplicadores de Lagrange, sendo equivalente a minimizar

$$f(\mathbf{u}) = \|\mathbf{T}\mathbf{u}\|_Y^2 - \sigma^2 \|\mathbf{u}\|_X^2,$$

sobre todo o espaço  $X$ . O máximo é obtido exatamente pela equação

$$\mathbf{T}^*\mathbf{T}\mathbf{u} = \sigma^2\mathbf{u},$$

nos dando um autovalor do operador de Gram  $\mathbf{T}^*\mathbf{T}$  e um autovetor unitário correspondente.

## 6. Aplicações

### 6.1. Cadeias de Markov.

### 6.2. Algoritmo de PageRank.

### 6.3. Decomposição ortogonal própria ou análise de componentes principais.

**6.4. Rigidez em equações diferenciais.** Em uma equação diferencial escalar de evolução da forma

$$\frac{dx}{dt} = \lambda x,$$

onde  $t$  representa uma variável temporal e  $x$  alguma quantidade de interesse, a solução é da forma

$$x(t) = x_0 e^{\lambda t}, \quad \forall t \in \mathbb{R}.$$

Note que o coeficiente  $\lambda$  tem dimensão de frequência, ou seja, de inversa multiplicativa da dimensão temporal. Em um problema desse tipo, uma escala de tempo típica é da ordem  $\mathcal{O}(1/|\lambda|)$ . Para uma boa resolução numérica da solução, especialmente via métodos explícitos, devemos ter um passo de tempo  $\Delta t$  da ordem de

$$\Delta t \ll \frac{1}{|\lambda|}.$$

Por outro lado, tipicamente estamos interessados em resolver o problema em uma escala de alguns *ciclos* típicos da equação, ou seja, buscamos resolver a equação em um intervalo temporal  $[0, T]$  com

$$T \sim \mathcal{O}(1/|\lambda|).$$

Numericamente, isso significa que  $T = n\Delta t$ , em  $n$  passos de tempo, com

$$n = \frac{T}{\Delta t} \gg 1.$$

Ou seja, perfeitamente razoável, aproximar a equação em  $n$  passos de tempo, bastando  $n \gg 1$ . Mesmo que queiramos resolver isso a longo prazo, com

$$T \gg \frac{1}{\lambda},$$

precisaríamos de um número de passos  $n$  satisfazendo

$$\sqrt{n} = \sqrt{\frac{T}{\Delta t}} \gg 1.$$

O problema acontece em sistemas de equações diferenciais, contendo quantidades diversas e escalas de tempo diferentes. Mais precisamente, considere um sistema de equações diferenciais lineares da forma

$$\frac{d\mathbf{u}}{dt} = \mathbf{A}\mathbf{u},$$

o papel do coeficiente  $\lambda$  acima é feito pelos autovalores de  $\mathbf{A}$ , no caso de  $\mathbf{A}$  ser diagonalizável, ou, mais geralmente, pelos valores singulares de  $\mathbf{A}$ . Eles são responsáveis por ditar as maiores e menores escalas de tempo do sistema.

Considerando o maior valor singular  $\sigma_1$  e o menor valor singular  $\sigma_n$ , o maior valor singular determina a escala de tempo mais rápida a ser resolvida e, assim, dita a resolução da malha temporal,

$$\Delta t \ll \frac{1}{\sigma_1},$$

enquanto que o menor valor singular dita as escalas mais lentas e, com isso, as grandes escalas de tempo, de interesse,

$$T \sim \frac{1}{\sigma_n}.$$

O custo computacional fica da ordem de

$$\frac{T}{\Delta t} \gg \frac{\sigma_1}{\sigma_n} = \kappa(\mathbf{A}).$$

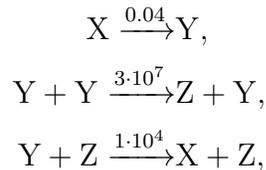
Ou seja, o número de condicionamento dita o custo computacional. Um número alto de condicionamento significa um grande custo computacional.

A ideia acima pode ser estendida a sistemas de equações não lineares,

$$\frac{d\mathbf{u}}{dt} = \mathbf{f}(t, \mathbf{u}),$$

onde as escalas de tempo são ditadas pelos valores singulares da diferencial  $D\mathbf{f}(t, \mathbf{u})$  do termo não linear.

Um exemplo clássico de equação diferencial rígida é o da reação química de Robertson, envolvendo três elementos químicos com constantes de reação bem distintas, a saber



gerando o sistema de equações diferenciais

$$\begin{cases} \dot{x} = -0.04x + 10^4 yz, \\ \dot{y} = 0.04x - 10^4 yz - 3 \times 10^7 y^2, \\ \dot{z} = 3 \times 10^7 y^2. \end{cases}$$

Observe a grande disparidade na ordem de grandeza dos coeficientes, levando a números de condicionamento exorbitantes. Métodos numéricos explícitos são impraticáveis. Métodos implícitos são essenciais na resolução desses sistemas, por serem capazes de dar passos de tempo muito maiores. Sistemas assim aparecem comumente em sistemas complexos de reações químicas e em diversos outros modelos heterogêneos.

OBSERVAÇÃO 6.1. Uma leitura um pouco mais precisa das relações  $a \ll b$  e  $c \sim b$  é a de que temos dada uma constante adimensional  $\delta > 0$ , representando alguma ordem de grandeza de interesse, digamos  $\delta = 1/10^k$ , para algum  $k$  fixo ao longo das relações (e.g.  $\delta = 1/100$ ), de forma que

$$\frac{a}{b} < \delta, \quad \left|1 - \frac{c}{b}\right| < \delta.$$

Há uma “folga” que nos permite manter essas relações  $a \ll b$  mesmo quando  $a/b < C\delta$ , para uma constante adimensional  $C$  próxima de 1. Com isso, obtemos

$$\frac{c}{a} = \frac{cb}{ba} \geq \frac{1-\delta}{\delta} \approx \frac{1}{\delta},$$

ou seja  $c \gg a$ . No caso em que  $c \gg b$ , então

$$\frac{c}{a} = \frac{cb}{ba} > \frac{1}{\delta^2},$$

por isso escrevemos

$$\sqrt{\frac{c}{a}} \gg 1.$$

Mais explicitamente, se, por exemplo,

$$\frac{a}{b} < \frac{1}{100}, \quad \frac{c}{b} > 100,$$

então

$$\frac{c}{a} > \frac{1}{10000},$$

ou seja,

$$\sqrt{\frac{c}{a}} > \frac{1}{100}.$$

## 7. Exercícios

### Exercícios

- 7.1. Verifique cada caracterização, no [Exemplo 1.1](#), da matriz  $\mathbf{A} \in \mathbb{R}^{2 \times 2}$ , para que seja autoadjunta, anti-autoadjunto e normal.
- 7.2. Verifique que o operador  $T_1$  definido na [Observação 2.3](#), no espaço quociente  $X/S_1$ , é um operador autoadjunto.
- 7.3. Encontre a decomposição em valores singulares da matriz

$$\mathbf{A} = \begin{bmatrix} 9 & -4 \\ 12 & 3 \end{bmatrix}.$$

- 7.4.** Sejam  $X$  e  $Y$  espaços vetoriais reais munidos de produto interno e seja  $T \in \mathcal{L}(X, Y)$ . Suponha que  $X$  e  $Y$  sejam de dimensão finita, com  $n = \dim(X)$ ,  $m = \dim(Y) \in \mathbb{N}$ . Considere uma decomposição em valores singulares de  $T$ , dada por bases  $\alpha = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  de  $X$  e  $\beta = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$  de  $Y$  e valores singulares  $\sigma_1 \geq \dots \geq \sigma_n \geq 0$ . Mostre que
- $((T\mathbf{v}_j, T\mathbf{v}_i)) = \sigma_j^2 \delta_{ij}$ , para  $i, j = 1, \dots, \min\{n, m\}$ .
  - $T\mathbf{v}_j \perp \text{span}\{\mathbf{v}_i, i \neq j\}$ , para  $j = 1, \dots, n$ .
  - $T^*T\mathbf{v}_j = \sigma_j^2 \mathbf{v}_j$ ,  $j = 1, \dots, \min\{n, m\}$ .
  - $\sigma_j^2$  é autovalor de  $T^*T$  com autovetor associado  $\mathbf{v}_j$ .
- 7.5.** Seja  $X$  um espaço vetorial real com produto interno e com dimensão  $n = \dim(X) \in \mathbb{N}$ . Seja  $T \in \mathcal{L}(X)$  um operador autoadjunto em  $X$ . Considere uma decomposição em valores singulares de  $T$ , dada por bases  $\alpha = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  de  $X$  e  $\beta = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$  de  $Y$  e valores singulares  $\sigma_1 \geq \dots \geq \sigma_n \geq 0$ . Mostre que cada  $\mathbf{v}_j$  é um autovetor de  $T$ , associado a um autovalor  $\lambda_j$  com  $\sigma_j = |\lambda_j|$ . Além disso,  $\mathbf{u}_j = \text{sgn}(\lambda_j)\mathbf{v}_j$ .
- 7.6.** Seja  $X$  um espaço vetorial real com produto interno e dimensão finita. Seja  $T \in \mathcal{L}(X)$  um operador anti-autoadjunto em  $X$ . Associe as bases e os valores singulares de uma decomposição em valores singulares de  $T$  à decomposição espectral dada pelo [Teorema 3.3](#).
- 7.7.** Seja  $X$  um espaço vetorial real com produto interno e dimensão finita. Seja  $T \in \mathcal{L}(X)$  um operador normal em  $X$ . Associe as bases e os valores singulares de uma decomposição em valores singulares de  $T$  à decomposição espectral dada pelo [Teorema 4.1](#).
- 7.8.** Sejam  $X$  e  $Y$  espaços vetoriais reais munidos de produto interno e seja  $T \in \mathcal{L}(X, Y)$ . Suponha que  $X$  e  $Y$  sejam de dimensão finita, com  $n = \dim(X)$ ,  $m = \dim(Y) \in \mathbb{N}$ . Considere uma decomposição em valores singulares de  $T$ , dada por bases  $\alpha = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  de  $X$  e  $\beta = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$  de  $Y$  e valores singulares  $\sigma_1 \geq \dots \geq \sigma_n \geq 0$ . Mostre que
- $((T^*\mathbf{u}_j, \mathbf{v}_i)) = \sigma_j \delta_{ij}$ , para  $i, j = 1, \dots, \min\{n, m\}$ .
  - $\mathbf{u}_j \in \ker(T^*)$ , para  $j = n + 1, \dots, m$ , caso  $n < m$ .
  - $\mathbf{v}_j \perp \text{Im}(T^*)$ , para  $j = m + 1, \dots, n$ , caso  $m < n$ .
  - Conclua que a decomposição em valores singulares de  $T^*$  é a mesma de  $T$ , com os mesmos valores singulares e com as bases apenas “trocando de papel” domínio/contra-domínio.
- 7.9.** Sejam  $X$  e  $Y$  espaços vetoriais reais munidos de produto interno e seja  $T \in \mathcal{L}(X, Y)$ . Suponha que  $X$  e  $Y$  sejam de dimensão finita, com  $n = \dim(X)$ ,  $m = \dim(Y) \in \mathbb{N}$ . Considere uma decomposição em valores singulares de  $T$ , dada por bases  $\alpha = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  de  $X$  e  $\beta = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$  de  $Y$  e

valores singulares  $\sigma_1 \geq \dots \geq \sigma_n \geq 0$ . Mostre que, para  $j = 1, \dots, \min\{n, m\}$ , os valores singulares  $\sigma_j$  são autovalores, também, de  $TT^* \in \mathcal{L}(Y)$ , com autovetores associados  $\mathbf{u}_j$ .

## Teoria Espectral em Espaços Vetoriais Complexos com Produto Interno

Variáveis complexas são fundamentais não apenas em aplicações, como em mecânica quântica e eletrônica, para citar algumas, mas também como ferramenta fundamental em análise e álgebra linear. Veremos, no que se segue, como ela facilita a análise espectral. Ela também é fundamental na análise de perturbação do espectro de um operador, conforme comentado na [Observação 5.1](#) e visto em mais detalhes na [Seção 6.1](#).

### 1. Produto interno em espaços vetoriais complexos

**1.1. Definição e exemplos.** Enquanto que nos reais, o valor absoluto  $|r|$  de um número real pode ser obtido via raiz quadrada do quadrado do número, i.e.  $|r| = \sqrt{r^2}$ , o valor absoluto de um número complexo  $z = a + ib$  é obtido tomando-se a raiz quadrada do produto entre o número e o seu conjugado,

$$|z| = \sqrt{z\bar{z}} = \sqrt{(a + ib)(a - ib)} = \sqrt{a^2 + b^2}.$$

Essa diferença nos obriga a modificar a [Definição 2.1](#) de produto escalar, de forma a ter um produto escalar positivo definido entre os complexos e recuperar uma norma através da raiz quadrada do produto escalar (veja [Exemplo 1.1](#)). Isso nos leva à seguinte definição mais geral de produto interno.

**DEFINIÇÃO 1.1 (Produto interno).** *Um **produto interno (hermitiano)** em um espaço vetorial complexo  $X$  é uma forma sesqui-linear  $((\cdot, \cdot)) : X \times X \rightarrow \mathbb{C}$  conjugada-simétrica e positiva definida, i.e.*

- (i)  $((\mathbf{u} + \lambda\mathbf{v}, \mathbf{w})) = ((\mathbf{u}, \mathbf{w})) + \lambda((\mathbf{v}, \mathbf{w}))$  e  $((\mathbf{u}, \mathbf{v} + \lambda\mathbf{w})) = ((\mathbf{u}, \mathbf{v})) + \bar{\lambda}((\mathbf{u}, \mathbf{w}))$ , para todo  $\mathbf{u}, \mathbf{v}, \mathbf{w} \in X$  e todo  $\lambda \in \mathbb{K}$ ;
- (ii)  $((\mathbf{u}, \mathbf{v})) = \overline{((\mathbf{v}, \mathbf{u}))}$ , para todo  $\mathbf{u}, \mathbf{v} \in X$ ; e
- (iii)  $((\mathbf{u}, \mathbf{u})) \geq 0$  real, para todo  $\mathbf{u} \in X$ , com  $((\mathbf{u}, \mathbf{u})) = 0$  se, e somente se,  $\mathbf{u} = \mathbf{0}$ .

OBSERVAÇÃO 1.1. Considerando a parte real de vetores complexos, podemos identificar o  $\mathbb{R}^n$  como um subespaço de  $\mathbb{C}^n$  e podemos considerar o produto escalar complexo restrito a esse subespaço de vetores reais. Nesse sentido, observe que a restrição do produto escalar dado na Definição 1.1 não está em contradição com o dado na Definição 2.1, visto que o conjugado de um número real é o próprio número real. Então, de fato, o produto interno dado na Definição 1.1 é uma *extensão* do produto interno real, dado na Definição 2.1.

EXEMPLO 1.1 (Produto escalar hermitiano em  $\mathbb{C}^n$ ). Em  $\mathbb{C}^n$ , o produto escalar complexo, também chamado produto escalar hermitiano, toma a forma

$$\mathbf{u} \cdot \mathbf{v} = z_1 \bar{w}_1 + \cdots + z_n \bar{w}_n = \sum_{j=1}^n z_j \bar{w}_j,$$

para todo  $\mathbf{u} = (z_1, \dots, z_n)$ ,  $\mathbf{v} = (w_1, \dots, w_n) \in \mathbb{C}^n$ . A norma associada ao produto escalar complexo é a norma  $\ell^2$  vista no Exemplo 1.1,

$$\|\mathbf{u}\| = \sqrt{|z_1|^2 + \cdots + |z_n|^2}.$$

EXEMPLO 1.2 (Produto interno em  $\ell^2(\mathbb{C})$ ). No caso de seqüências de números complexos, consideramos

$$\ell^2(\mathbb{C}) = \{z = (z_n)_{n \in \mathbb{N}} \in \mathbb{C}^{\mathbb{N}}; \sum_{n \in \mathbb{N}} |z_n|^2 < \infty\},$$

e o produto interno tem a forma

$$((z, w)) = \sum_{n \in \mathbb{N}} z_n \bar{w}_n,$$

para  $z = (z_n)_{n \in \mathbb{N}} \in \ell^2(\mathbb{C})$  e  $w = (w_n)_{n \in \mathbb{N}} \in \ell^2(\mathbb{C})$ .

OBSERVAÇÃO 1.2. A versão complexa  $\ell^2_\varphi(\mathbb{C})$  de  $\ell^2$  com peso, a função peso  $\varphi$  continua sendo uma seqüência de números *reais* positivos.

**1.2. Norma proveniente do produto interno complexo.** A principal razão da definição de produto interno usando o conjugado de um número complexo é para recuperarmos a norma  $\ell^2$ ,

$$\|\mathbf{u}\| = \sqrt{((\mathbf{u}, \mathbf{v}))}.$$

Mas observe que, enquanto  $((\mathbf{u}, \mathbf{u}))$  é sempre real, para qualquer vetor complexo  $\mathbf{u}$ , visto que, trocando a ordem dos vetores, obtemos que esse produto interno é igual ao seu conjugado,  $((\mathbf{u}, \mathbf{u})) = \overline{((\mathbf{u}, \mathbf{u}))}$ , no caso de dois vetores diferentes, o produto interno  $((\mathbf{u}, \mathbf{v}))$  pode ser complexo. Associado a isso, a identidade (2.1) fica ligeiramente modificada.

TEOREMA 1.1. *Dado um espaço vetorial complexo  $X$  munido de um produto interno  $((\cdot, \cdot))$  e norma associada  $\|\cdot\|$ , vale a identidade*

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + 2\operatorname{Re}((\mathbf{u}, \mathbf{v})) + \|\mathbf{v}\|^2. \quad (1.1)$$

A desigualdade de Cauchy-Schwarz (2.2) e a desigualdade triangular (2.3) permanecem idênticas.

TEOREMA 1.2 (Desigualdade de Cauchy-Schwarz nos complexos). *Dado um espaço vetorial complexo  $X$  munido de um produto interno  $((\cdot, \cdot))$ , com norma associada  $\|\cdot\|$ , vale a desigualdade*

$$|((\mathbf{u}, \mathbf{v}))| \leq \|\mathbf{u}\|\|\mathbf{v}\|, \quad \forall \mathbf{u}, \mathbf{v} \in X. \quad (1.2)$$

TEOREMA 1.3 (Desigualdade triangular nos complexos). *Dado um espaço vetorial complexo  $X$  munido de um produto interno  $((\cdot, \cdot))$ , a norma  $\|\cdot\|$  associada ao produto interno satisfaz a desigualdade triangular*

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|, \quad \forall \mathbf{u}, \mathbf{v} \in X. \quad (1.3)$$

Com isso, vemos que

$$\|\mathbf{u}\| = \sqrt{((\mathbf{u}, \mathbf{u}))}$$

também define uma norma em um espaço vetorial complexo com produto interno, assim como visto na [Observação 2.1](#) no caso real.

OBSERVAÇÃO 1.3 (Não existência de produto interno bilinear nos complexos). A definição de produto interno nos complexos via forma conjugada-simétrica não é uma mera modificação para termos propriedades melhores. Ela é essencial. É claro que existem formas bilineares simétricas nos complexos, mas pode-se mostrar que não existe forma bilinear simétrica que seja positiva definida (veja [Exercício 7.1](#)).

**1.3. Ortogonalidade.** A definição de ortogonalidade nos complexos é a mesma que nos reais.

DEFINIÇÃO 1.2 (Ortogonalidade nos complexos). *Seja  $X$  um espaço vetorial complexo com produto interno  $((\cdot, \cdot))$ . Dois vetores  $\mathbf{u}, \mathbf{v}$  em um espaço vetorial real  $X$  munido de um produto interno  $((\cdot, \cdot))$  são ditos **ortogonais** em relação a esse produto interno, quando*

$$((\mathbf{u}, \mathbf{v})) = 0.$$

Nesse caso, escrevemos  $\mathbf{u} \perp \mathbf{v}$ . Um vetor  $\mathbf{u}$  é dito ortogonal a um subconjunto  $A \subset X$  quando  $\mathbf{u} \perp \mathbf{v}$  para todo  $\mathbf{v} \in A$ . Nesse caso, escrevemos  $\mathbf{u} \perp A$ . Mais geralmente, dois subconjuntos  $A$  e  $B$  de  $X$  são ditos ortogonais entre si quando  $\mathbf{u} \perp \mathbf{v}$  para quaisquer

$\mathbf{u} \in A$  e  $\mathbf{v} \in B$ . Nesse caso, escrevemos  $A \perp B$ . Dado um subconjunto  $A \in X$ , denotamos o seu perpendicular por

$$A^\perp = \{\mathbf{u} \in X; \mathbf{u} \perp A\}.$$

OBSERVAÇÃO 1.4. A analogia entre os números complexos  $\mathbb{C}$  e o plano  $\mathbb{R}^2$  é útil em vários aspectos. Por exemplo, multiplicar o número complexo  $z = a + ib$  por  $i$  é como uma rotação no plano de  $90^\circ$ ,  $iz = -b + ia$ . Mas essa analogia não deve ser extrapolada. No plano  $\mathbb{R}^2$ , os vetores  $(a, b)$  e  $(-b, a)$  são, de fato, ortogonais, mas, nos complexos, os escalares  $z = a + ib$  e  $iz = -b + ia$  não são ortogonais. O produto interno entre eles é

$$\begin{aligned} z \cdot iz &= z\overline{(iz)} = (a + ib)\overline{(-b + ia)} = (a + ib)(-b - ia) = -ab - ia^2 - ib^2 - i^2ab \\ &= -ab - i(a^2 + b^2) + ab = -i(a^2 + b^2) \neq 0, \end{aligned}$$

para  $z = a + ib \neq 0$ . Em  $\mathbb{C}$ , os elementos  $z$  e  $iz$  são meros escalares, não vetores, e um é um múltiplo escalar do outro. Ou seja, estão no mesmo unidimensional. O espaço  $\mathbb{C} = \mathbb{C}^1$  é um espaço unidimensional. Nenhum vetor vai ser ortogonal a outro.

OBSERVAÇÃO 1.5. Complementando a [Observação 1.4](#), precisamos de pelo menos duas dimensões para termos vetores ortogonais. Por exemplo, os vetores  $\mathbf{u} = (1, i)$  e  $\mathbf{v} = (1, -i)$  são, de fato, ortogonais, visto que

$$\mathbf{u} \cdot \mathbf{v} = (1, i) \cdot (1, -i) = 1 \times \bar{1} + i \times \overline{-i} = 1 + i \times i = 1 + i^2 = 1 - 1 = 0.$$

Assim como no caso real, o perpendicular  $A^\perp$  de um conjunto qualquer  $A$  é sempre um subespaço linear. Caso  $A$  também seja subespaço, então os dois formam uma soma direta que é igual ao espaço  $X$ .

TEOREMA 1.4. *Seja  $X$  um espaço vetorial complexo com produto interno  $((\cdot, \cdot))$ . Dado um subconjunto  $A \in X$ , o seu perpendicular  $A^\perp$  é um subespaço vetorial de  $X$ . Se  $S$  é um subespaço vetorial de  $X$ , então  $S$  e  $S^\perp$  são linearmente independentes.*

Com o conceito de ortogonalidade, podemos falar de projeção ortogonal nos complexos, também.

DEFINIÇÃO 1.3 (Projeção ortogonal nos complexos). *Seja  $X$  um espaço vetorial complexo com produto interno  $((\cdot, \cdot))$ . Uma projeção  $\mathbf{P} : X \rightarrow X$  sobre um subespaço  $S$  é dita uma **projeção ortogonal** quando  $\ker \mathbf{P} = S^\perp$ , ou seja, quando  $\mathbf{P}$  é uma projeção sobre  $S$  ao longo do subespaço perpendicular  $S^\perp$ .*

A projeção ortogonal na direção de um vetor é obtida da mesma forma que nos reais. A demonstração é idêntica, observando que usamos a linearidade apenas no primeiro argumento do produto interno.

TEOREMA 1.5 (Projeção ortogonal na direção de um vetor). *Seja  $X$  um espaço vetorial complexo com produto interno  $((\cdot, \cdot))$  e norma associada  $\|\cdot\|$ . Seja  $\mathbf{v} \in X$  um vetor não nulo. Então*

$$P_{\mathbf{v}}\mathbf{u} = \frac{((\mathbf{u}, \mathbf{v}))}{\|\mathbf{v}\|^2}\mathbf{v} \quad (1.4)$$

*defina um operador de projeção  $P_{\mathbf{v}} : X \rightarrow X$  sobre o subespaço  $S = \text{span}\{\mathbf{v}\}$  gerado por  $\mathbf{v}$ , ao longo do subespaço ortogonal  $S^\perp$ .*

OBSERVAÇÃO 1.6. É importante ressaltar que, agora, a ordem dos vetores que aparece no operador projeção (1.4) é importante. Ou seja, ao projetarmos um vetor  $\mathbf{u}$  na direção de um vetor  $\mathbf{v}$ , é importante definirmos a projeção usando o fator multiplicativo  $((\mathbf{u}, \mathbf{v}))/\|\mathbf{v}\|^2$ , ao passo que, nos reais, a ordem  $((\mathbf{u}, \mathbf{v}))$  ou  $((\mathbf{v}, \mathbf{u}))$  não faz diferença.

Como a projeção ortogonal nos complexos é igual a nos reais, o processo de Gram-Schmidt também é o mesmo.

TEOREMA 1.6 (Processo de ortogonalização de Gram-Schmidt nos complexos). *Seja  $X$  um espaço vetorial complexo com produto interno  $((\cdot, \cdot))$  e norma associada  $\|\cdot\|$ . Dada uma sequência  $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$  de vetores linearmente independentes, a sequência  $\{\mathbf{w}_1, \dots, \mathbf{w}_m\}$  dada por*

$$\mathbf{w}_j = \frac{\mathbf{w}'_j}{\|\mathbf{w}'_j\|}, \quad \mathbf{w}'_j = \mathbf{v}_j - P_{\mathbf{w}_1}\mathbf{v}_j - \dots - P_{\mathbf{w}_{j-1}}\mathbf{v}_j,$$

*para  $j = 1, \dots, m$ , define vetores ortogonais entre si, com norma unitária e gerando os mesmos subespaços, i.e.*

$$\|\mathbf{w}_j\| = 1, \quad ((\mathbf{w}_i, \mathbf{w}_j)) = 0, \quad \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_j\} = \text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_j\},$$

*para todo  $i, j = 1, \dots, m$ , com  $i \neq j$ .*

Assim como nos reais, podemos usar o processo de Gram-Schmidt para construir a projeção ortogonal sobre subespaços vetoriais de dimensão finita.

TEOREMA 1.7. *Seja  $X$  um espaço vetorial real com produto interno  $((\cdot, \cdot))$  e seja  $S \subset X$  um subespaço vetorial de  $X$  de dimensão finita. Então existe uma projeção ortogonal  $P_S : X \rightarrow X$  sobre  $S$  e vale*

$$X = S \oplus S^\perp.$$

OBSERVAÇÃO 1.7. Assim como nos reais, o resultado do Teorema 1.7 pode ser usado para demonstrar a existência de projeção ortogonal sobre subespaços de codimensão finita. Veja Exercício 6.2.

Com projeções ortogonais, temos uma generalização do Teorema de Pitágoras.

TEOREMA 1.8. *Seja  $X$  um espaço vetorial complexo com produto interno  $((\cdot, \cdot))$  e suponha que  $P_S$  seja uma projeção ortogonal sobre um subespaço  $S$ . Então  $P_{S^\perp} = I - P_S$  é a projeção ortogonal sobre  $S^\perp$  e vale a identidade*

$$\|\mathbf{u}\|^2 = \|P_S \mathbf{u}\|^2 + \|P_{S^\perp} \mathbf{u}\|^2, \quad \forall \mathbf{u} \in X. \quad (1.5)$$

Temos, da mesma forma, a seguinte caracterização.

TEOREMA 1.9. *Seja  $X$  um espaço vetorial complexo com produto interno  $((\cdot, \cdot))$  e suponha que  $P$  seja uma projeção em  $X$ . Então  $P$  é projeção ortogonal se, e somente se,*

$$((P\mathbf{u}, Q\mathbf{u})) = 0,$$

para todo  $\mathbf{u} \in X$ , onde  $Q = I - P$ .

A projeção ortogonal complexa também tem a propriedade de minimizar a distância até um subespaço vetorial.

TEOREMA 1.10. *Seja  $X$  um espaço vetorial complexo com produto interno  $((\cdot, \cdot))$  e norma associada  $\|\cdot\|$ . Seja  $S \subset X$  e suponha que  $P_S$  seja a projeção ortogonal sobre  $S$ . Então*

$$\|\mathbf{u} - P_S \mathbf{u}\| \leq \|\mathbf{u} - \mathbf{v}\|, \quad \forall \mathbf{v} \in S.$$

Também podemos falar de base ortonormal nos complexos.

DEFINIÇÃO 1.4. *Seja  $X$  um espaço vetorial complexo com produto interno  $((\cdot, \cdot))$  e norma associada  $\|\cdot\|$  e dimensão finita  $n = \dim(X) < \infty$ . Uma **base ortonormal** de  $X$  é uma base  $\mathcal{B} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  tal que cada vetor tem norma unitária e os vetores são ortogonais entre si, ou seja,*

$$\|\mathbf{w}_j\| = 1, \quad ((\mathbf{w}_i, \mathbf{w}_j)) = 0, \quad i, j = 1, \dots, n, \quad i \neq j.$$

O processo de ortogonalização de Gram-Schmidt nos garante a existência de uma base ortonormal de qualquer subespaço de dimensão finita, também nos complexos.

TEOREMA 1.11. *Seja  $X$  um espaço vetorial complexo com produto interno e seja  $S \subset X$  um subespaço de  $X$  não trivial de dimensão finita. Então  $S$  possui uma base ortonormal.*

**1.4. Realização do dual e da adjunta nos complexos.** Assim como no caso real, funcionais lineares em espaços vetoriais complexos com produto interno também podem ser representados por vetores no próprio espaço.

TEOREMA 1.12 (Teorema da Representação de Riesz em dimensão finita no caso complexo). *Seja  $\mathfrak{f} \in X^*$  um funcional linear em um espaço vetorial complexo  $X$  de dimensão finita, munido de um produto interno  $((\cdot, \cdot))$ . Então existe um único  $\mathbf{v} \in X$  tal que*

$$\langle \mathfrak{f}, \mathbf{u} \rangle = ((\mathbf{u}, \mathbf{v})),$$

para todo  $\mathbf{u} \in X$ .

DEMONSTRAÇÃO. A demonstração é essencialmente a mesma, com pequenas modificações que deixamos a cargo do leitor.  $\square$

OBSERVAÇÃO 1.8. Com a forma sesquilinear do produto interno hermitiano, o morfismo  $J : X \rightarrow X^*$  definido por

$$\langle J(\mathbf{v}), \mathbf{u} \rangle_{X^*, X} = ((\mathbf{u}, \mathbf{v})),$$

para  $\mathbf{u}, \mathbf{v} \in X$ , que é uma bijeção, graças ao Teorema 1.12, não é, no entanto, uma transformação linear. Ela é uma transformação *anti-linear*, satisfazendo

$$J(\mathbf{v} + \zeta \mathbf{w}) = J(\mathbf{v}) + \bar{\zeta} J(\mathbf{w}),$$

para todo  $\mathbf{v}, \mathbf{w} \in X$ ,  $\zeta \in \mathbb{C}$ . De fato, isso segue da própria definição de  $J$ , da propriedade sesquilinear da forma e da álgebra (adição e multiplicação por escalar) definida em  $X^*$ :

$$\begin{aligned} \langle J(\mathbf{v} + \zeta \mathbf{w}), \mathbf{u} \rangle_{X^*, X} &= ((\mathbf{u}, \mathbf{v} + \zeta \mathbf{w})) \\ &= ((\mathbf{u}, \mathbf{v})) + \bar{\zeta} ((\mathbf{u}, \mathbf{w})) \\ &= \langle J(\mathbf{v}), \mathbf{u} \rangle_{X^*, X} + \bar{\zeta} \langle J(\mathbf{w}), \mathbf{u} \rangle_{X^*, X} \\ &= \langle J(\mathbf{v}) + \bar{\zeta} J(\mathbf{w}), \mathbf{u} \rangle_{X^*, X}. \end{aligned}$$

Apesar de não ser linear,  $J$  ainda assim é uma bijeção. Em duais de espaços complexos, este tipo de morfismo via bijeções anti-lineares é uma forma natural de *isomorfismo*.

Uma vez tendo realizações dos duais, podemos interpretar a transformação adjunta  $\mathbb{T}^* \in \mathcal{L}(Y^*, X^*)$ , discutida na Seção 4, como uma transformação em  $\mathcal{L}(Y, X)$ , também nos complexos.

DEFINIÇÃO 1.5 (Transformação adjunta em espaços com produto interno complexos). *Sejam  $X$  e  $Y$  espaços vetoriais complexos com produto interno e sejam  $((\cdot, \cdot))_X$  e  $((\cdot, \cdot))_Y$  os respectivos produtos internos. Seja  $\mathbb{T} \in \mathcal{L}(X, Y)$  uma transformação linear de  $X$  em  $Y$ . Então a **transformação adjunta**  $\mathbb{T}^* : Y \rightarrow X$  de  $\mathbb{T}$ , é definida por*

$$((\mathbb{T}\mathbf{u}, \mathbf{v}))_Y = ((\mathbf{u}, \mathbb{T}^*\mathbf{v}))_X,$$

para todo  $\mathbf{u} \in X$ ,  $\mathbf{v} \in Y$ , e é uma representação, via produto interno, da transformação adjunta de  $Y^*$  em  $X^*$  vista na [Seção 4](#) (veja [Observação 1.9](#)).

OBSERVAÇÃO 1.9. A transformação adjunta  $\mathbb{T}^* \in \mathcal{L}(Y, X)$  vista na [Definição 1.5](#) é uma representação da adjunta vista na [Seção 4](#), que definimos temporariamente por  $\tilde{\mathbb{T}}^* \in \mathcal{L}(Y^*, X^*)$ , no sentido de que

$$\langle \tilde{\mathbb{T}}^* \mathbf{g}, \mathbf{u} \rangle_{X^*, X} = ((\mathbf{u}, \mathbb{T}^* \mathbf{J}(\mathbf{g})))_X = ((\mathbb{T} \mathbf{u}, \mathbf{J}(\mathbf{g})))_Y = \langle \mathbf{g}, \mathbb{T} \mathbf{u} \rangle_{Y^*, Y},$$

para todo  $\mathbf{u} \in X$  e todo  $\mathbf{g} \in Y^*$ , onde  $\mathbf{J} \in \mathcal{L}(Y^*, Y)$  é o morfismo (anti-linear) do dual  $Y^*$  em  $Y$  discutido na [Observação 1.8](#).

Com essa representação via produto interno complexo, que é uma forma sesqui-linear, as propriedades elencadas na [Proposição 4.2](#) sofrem uma pequena modificação, nos dando a relação

$$(\lambda \mathbb{T})^* = \bar{\lambda} \mathbb{T}^*,$$

com o conjugado de  $\lambda$ , ao invés do próprio  $\lambda$ .

PROPOSIÇÃO 1.1. *Dados espaços vetoriais  $X$ ,  $Y$  e  $Z$  complexos,  $\lambda \in \mathbb{C}$  e transformações lineares  $\mathbb{T}, \mathbb{L} \in \mathcal{L}(X, Y)$ ,  $\mathbb{S} \in \mathcal{L}(Y, Z)$ , então*

$$(\mathbb{T} + \mathbb{L})^* = \mathbb{T}^* + \mathbb{L}^*, \quad (\lambda \mathbb{T})^* = \bar{\lambda} \mathbb{T}^*, \quad (\mathbb{S} \circ \mathbb{T})^* = \mathbb{T}^* \circ \mathbb{S}^*.$$

Além disso, caso  $\mathbb{T}$  seja invertível, então

$$(\mathbb{T}^{-1})^* = (\mathbb{T}^*)^{-1}.$$

A representação matricial da adjunto também tem uma pequena modificação.

TEOREMA 1.13. *Sejam  $X$  e  $Y$  espaços vetoriais complexos de dimensão finita  $n = \dim(X) \in \mathbb{N}$  e  $m = \dim(Y) \in \mathbb{N}$ . Seja  $\mathbb{T} \in \mathcal{L}(X, Y)$  uma transformação linear de  $X$  em  $Y$ . Dadas bases  $\alpha$ , de  $X$ , e  $\beta$ , de  $Y$ , sejam  $\alpha'$  e  $\beta'$  as respectivas bases duais. A representação matricial  $[\mathbb{T}^*]_{\alpha'}^{\beta'}$ , da transformação adjunta é a transposta conjugada da representação matricial  $[\mathbb{T}]_{\beta}^{\alpha}$ , i.e. se*

$$[\mathbb{T}]_{\beta}^{\alpha} = (a_{ij})_{i,j=1}^{i=m, j=n},$$

então

$$[\mathbb{T}^*]_{\alpha'}^{\beta'} = (\bar{a}_{ji})_{i,j=1}^{i=n, j=m}.$$

EXEMPLO 1.3. Considerando a matriz  $\mathbf{A} \in \mathbb{C}^{2 \times 3}$  dada por

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

a sua conjugada  $\mathbf{A}^* \in \mathbb{C}^{3 \times 2}$  toma a forma

$$\mathbf{A}^* = \begin{bmatrix} 1 - i & -3i \\ 2 & 4 \\ i & 2 + i \end{bmatrix}.$$

OBSERVAÇÃO 1.10. A representação vista no [Teorema 4.1](#), onde a adjunta é meramente a transposta, sem o conjugado, é válida também nos espaços complexos. Não restringimos aquele resultado a espaços vetoriais reais. No entanto, aquela realização do dual é feita naquele contexto de formas bilineares. Ao representarmos essas transformações adjuntas via produto interno hermitiano, no entanto, essas representações se comportam de maneira diferente, como visto no [Exemplo 1.3](#).

**1.5. Operadores especiais.** Da mesma forma, alguns operadores tem propriedades especiais em relação a simetrias com a adjunta.

DEFINIÇÃO 1.6. *Seja  $X$  um espaço vetorial complexo com produto interno. Seja  $T \in \mathcal{L}(X)$  um operador em  $X$  e considere a sua adjunta  $T^* \in \mathcal{L}(X)$  relativa a esse produto interno.*

- (1) Dizemos que  $T$  é **hermitiano** quando  $T^* = T$ .
- (2) Dizemos que  $T$  é **anti-hermitiano** quando  $T^* = -T$ .
- (3) Dizemos que  $T$  é **normal** quando  $TT^* = T^*T$ .
- (4) Dizemos que  $T$  é **unitário** quando  $T^*T = TT^* = I$ , ou seja,  $T^* = T^{-1}$ .

Mesmo nos complexos, as projeções ortogonais são caracterizadas como sendo as projeções hermitianas.

TEOREMA 1.14. *Seja  $X$  um espaço vetorial complexo com produto interno e seja  $P \in \mathcal{L}(X)$  uma projeção. Então  $P$  é projeção ortogonal se, e somente se,  $P$  é hermitiana, i.e.  $P^* = P$ .*

Um operador unitário também é caracterizado por preservar a norma, como no caso dos operadores ortogonais, visto no [Teorema 1.2](#), e por preservar o produto interno entre vetores, analogamente ao visto na [Observação 1.2](#).

Um operador normal também é caracterizado como no [Teorema 1.4](#) e no [Teorema 1.5](#).

## 2. Decomposição de Schur e a triangularização ortogonal

Vimos, no [Capítulo 5](#), a importância da forma triangular, de um operador, no estudo dos seus autovalores, autoespaços, autoespaços generalizadas e na forma canônica de Jordan. Uma versão daquele resultado usando bases ortogonais é um resultado fundamental conhecido como *decomposição de Schur*.

**2.1. Triangularização via subespaços invariantes ortogonais encaixantes.** Seja  $T : X \rightarrow X$  um espaço vetorial complexo  $X$  de dimensão finita  $n = \dim(X) \in \mathbb{N}$ . Então existe uma base  $\mathcal{B}$  satisfazendo as condições do [Teorema 3.2](#) e, portanto,  $T$  possui a forma triangular superior nessa base.

**TEOREMA 2.1.** *Seja  $T : X \rightarrow X$  um operador linear em um espaço vetorial complexo  $X$  com produto interno e de dimensão finita  $n = \dim(X) \in \mathbb{N}$ . Então existe uma base ortonormal  $\mathcal{B} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  de  $X$  tal que*

$$T(\text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_j\}) \subset \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_j\},$$

para todo  $j = 1, \dots, n$ .

**DEMONSTRAÇÃO.** Podemos provar por indução. Em uma dimensão, não há nada o que demonstrar. Qualquer base formada por um vetor unitário basta. Suponha, então, o resultado válido até dimensão  $n-1$ . Vamos provar para dimensão  $n \in \mathbb{N}$ ,  $n \geq 2$ . Seja, então,  $X$  um espaço vetorial complexo com produto interno e de dimensão  $n \in \mathbb{N}$ . Seja  $T \in \mathcal{L}(X)$ . Graças ao [Teorema 3.1](#),  $T$  possui pelo menos um autovalor  $\lambda_1 \in \mathbb{C}$ . Seja  $\mathbf{w}_1 \in X$  um autovetor associado a esse autovalor. Podemos dividir pela sua norma, se necessário, e assumir que  $\mathbf{w}_1$  é unitário, i.e.

$$\|\mathbf{w}_1\| = 1.$$

Seja

$$P_1 \mathbf{u} = ((\mathbf{u}, \mathbf{w}_1)) \mathbf{w}_1$$

a projeção ortogonal sobre

$$S_1 = \text{span}\{\mathbf{w}_1\}.$$

Seja  $S_1^\perp$  o subespaço vetorial perpendicular de  $S_1$ , de modo que

$$X = S_1 \oplus S_1^\perp.$$

Considere o operador a restrição de  $T$  ao subespaço perpendicular  $S_1^\perp$  composto com a projeção sobre o próprio  $S_1^\perp$ , i.e.

$$T_1 = Q_1 T \Big|_{S_1^\perp},$$

onde

$$Q_1 = I - P_1.$$

Esse operador pode ser visto como um operador em  $\mathcal{L}(S_1^\perp)$ , que é um subespaço de dimensão  $n-1$ . Pela hipótese de indução, existe uma base ortonormal

$$\{\mathbf{w}_2, \dots, \mathbf{w}_n\}$$

de  $S_1^\perp$  tal que

$$T_1 \operatorname{span}\{\mathbf{w}_2, \dots, \mathbf{w}_j\} \subset \operatorname{span}\{\mathbf{w}_2, \dots, \mathbf{w}_j\},$$

para todo  $j = 2, \dots, n-1$ .

Como  $\{\mathbf{w}_2, \dots, \mathbf{w}_n\}$  é uma base ortonormal no espaço ortogonal a  $\mathbf{w}_1$ , então

$$\mathcal{B} = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n\}$$

é uma base ortonormal do espaço  $n$ -dimensional  $X$ . Além disso, dado

$$\mathbf{u} = x_1 \mathbf{w}_1 + \dots + x_j \mathbf{w}_j \in \operatorname{span}\{\mathbf{w}_1, \dots, \mathbf{w}_j\},$$

temos

$$\begin{aligned} T\mathbf{u} &= TP_1\mathbf{u} + TQ_1\mathbf{u} = T(x_1\mathbf{w}_1) + T_1(x_2\mathbf{w}_2 + \dots + x_j\mathbf{w}_j) \\ &= x_1\lambda_1\mathbf{w}_1 + T_1(x_2\mathbf{w}_2 + \dots + x_j\mathbf{w}_j). \end{aligned}$$

Como

$$x_1\lambda_1\mathbf{w}_1 \in \operatorname{span}\{\mathbf{w}_1\}$$

e

$$T_1(x_2\mathbf{w}_2 + \dots + x_j\mathbf{w}_j) \in \operatorname{span}\{\mathbf{w}_2, \dots, \mathbf{w}_j\},$$

deduzimos que

$$T\mathbf{u} \in \operatorname{span}\{\mathbf{w}_1, \dots, \mathbf{w}_j\},$$

para todo

$$\mathbf{u} \in \operatorname{span}\{\mathbf{w}_1, \dots, \mathbf{w}_j\},$$

para todo  $j = 2, \dots, n$ , concluindo a demonstração. □

**2.2. Decomposição.** Temos, também, a versão matricial

**TEOREMA 2.2.** *Seja  $\mathbf{A} \in \mathbb{C}^n$  uma matriz complexa  $n \times n$ ,  $n \in \mathbb{N}$ . Então existe uma matriz unitária  $\mathbf{U} \in \mathbb{C}^n$  e uma matriz triangular superior  $\mathbf{R} \in \mathbb{C}^n$  tal que*

$$\mathbf{R} = \mathbf{U}^* \mathbf{A} \mathbf{U}.$$

**DEMONSTRAÇÃO.** Basta aplicar o **Teorema 2.1** e formar a matriz  $\mathbf{U}$  com as colunas dadas pelas coordenadas dos vetores  $\mathbf{w}_1, \dots, \mathbf{w}_n$ . □

### 2.3. Propriedades e exemplos.

OBSERVAÇÃO 2.1. Uma propriedade importante da decomposição

$$\mathbf{R} = \mathbf{U}^* \mathbf{A} \mathbf{U}$$

vem do fato de  $\mathbf{U}$  ser unitária e, com isso, o seu número de condicionamento ser exatamente 1. Assim, os números de condicionamento de  $\mathbf{A}$  e  $\mathbf{R}$  coincidem, i.e.

$$\kappa(\mathbf{R}) = \kappa(\mathbf{A}),$$

visto que  $\kappa(\mathbf{U}) = 1$ . Verifique.

EXEMPLO 2.1. Considere a matriz

$$\mathbf{A} = \begin{bmatrix} -499 & 500 \\ -500 & 501 \end{bmatrix}.$$

O único autovalor é  $\lambda = 1$ , com autoespaço

$$V(\mathbf{A}, 1) = \{(x, y); y = x\}.$$

Um autovetor unitário é

$$\mathbf{w}_1 = \frac{\sqrt{2}}{2}(1, 1).$$

Para completar até uma base ortonormal, tomamos

$$\mathbf{w}_2 = \frac{\sqrt{2}}{2}(-1, 1).$$

Temos

$$\mathbf{A} \mathbf{w}_2 = \frac{\sqrt{2}}{2} \begin{bmatrix} -499 & 500 \\ -500 & 501 \end{bmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \frac{\sqrt{2}}{2} \begin{pmatrix} 999 \\ 1001 \end{pmatrix} = 1000 \mathbf{w}_1 + \mathbf{w}_2.$$

O operador unitário associado a essa base, que é matriz mudança de base, da base canônica  $e$  para a base  $\mathcal{B} = \{\mathbf{w}_1, \mathbf{w}_2\}$ , é

$$\mathbf{U} = \frac{\sqrt{2}}{2} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

A forma triangular na decomposição de Schur é

$$\mathbf{R} = \mathbf{U}^* \mathbf{A} \mathbf{U} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -499 & 500 \\ -500 & 501 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1000 \\ 0 & 1 \end{bmatrix}$$

A forma triangular também pode ser obtida diretamente de

$$\mathbf{A} \mathbf{w}_1 = \mathbf{w}_1, \quad \mathbf{A} \mathbf{w}_2 = 1000 \mathbf{w}_1 + \mathbf{w}_2.$$

OBSERVAÇÃO 2.2. Em comparação, fazendo a forma normal de Jordan da matriz do Exemplo 2.1, a saber

$$\mathbf{A} = \begin{bmatrix} -499 & 500 \\ -500 & 501 \end{bmatrix}.$$

A sua forma normal é

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

Para chegar na forma normal, podemos começar com qualquer vetor que não seja autovetor, ou seja, que não seja múltiplo de  $(1, 1)$ . Por exemplo,

$$\mathbf{w}_2 = (0, 1).$$

Em seguida, tomamos

$$\mathbf{w}_1 = \mathbf{A}\mathbf{w}_2 = (500, 501)$$

A matriz mudança de base é

$$\mathbf{M} = \begin{bmatrix} 500 & 0 \\ 501 & 1 \end{bmatrix}$$

Nesse caso, a inversa de  $\mathbf{M}$  é

$$\mathbf{M}^{-1} = \begin{bmatrix} 0.002 & 0 \\ -1.002 & 1 \end{bmatrix}$$

e temos  $\mathbf{J} = \mathbf{M}^{-1}\mathbf{A}\mathbf{M}$ , i.e.

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0.002 & 0 \\ -1.002 & 1 \end{bmatrix} \begin{bmatrix} -499 & 500 \\ -500 & 501 \end{bmatrix} \begin{bmatrix} 500 & 0 \\ 501 & 1 \end{bmatrix}.$$

O número de condicionamento de  $\mathbf{M}$  (assim como o de sua inversa) é

$$\kappa(\mathbf{M}) = 1002.$$

Verifique que

$$\kappa(\mathbf{R}) = \kappa(\mathbf{A}) = 10^6$$

enquanto que

$$\kappa(\mathbf{M})\kappa(\mathbf{J})\kappa(\mathbf{M}^{-1}) \sim 5 \times 10^6.$$

### 3. Diagonalização de operadores normais

Vimos, na [Definição 1.6](#), os operadores normais  $T \in \mathcal{L}(X)$ , em um espaço vetorial complexo  $X$  com produto interno, que são aqueles que comutam com a sua adjunta, i.e.

$$T^*T = TT^*.$$

Veremos, aqui, que essa propriedade caracteriza os operadores diagonalizáveis em bases ortonormais. Ou seja,  $T$  possui uma base ortonormal de autovetores se, e somente se,  $T$  é normal.

**3.1. Operadores diagonalizáveis em alguma base ortonormal são normais.** Vejamos a ida.

**TEOREMA 3.1.** *Seja  $T \in \mathcal{L}(X)$  um operador linear em um espaço vetorial complexo  $X$  com produto interno. Suponha que  $X$  tenha uma base ortonormal de autovetores de  $T$ . Então  $T$  é normal.*

**DEMONSTRAÇÃO.** Seja  $\mathfrak{b} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  uma base ortonormal de  $X$  composta de autovetores de  $T$ . Então, para cada  $j = 1, \dots, n$ , existe um complexo  $\lambda_j \in \mathbb{C}$  tal que

$$T\mathbf{w}_j = \lambda_j\mathbf{w}_j.$$

Nesse particular de base ortonormal, segue que  $\mathbf{w}_j$  também é autovetor da adjunta  $T^*$ , com autovalor  $\bar{\lambda}_j$  (isso não é verdade em geral, veja [Observação 3.1](#)). De fato, temos

$$\begin{aligned} ((T^*\mathbf{w}_j, \mathbf{w}_i)) &= \overline{((\mathbf{w}_i, T^*\mathbf{w}_j))} = \overline{((T\mathbf{w}_i, \mathbf{w}_j))} = \overline{((\lambda_i\mathbf{w}_i, \mathbf{w}_j))} \\ &= ((\mathbf{w}_j, \lambda_i\mathbf{w}_i)) = \bar{\lambda}_i((\mathbf{w}_j, \mathbf{w}_i)). \end{aligned}$$

Caso  $i \neq j$ , obtemos que

$$((T^*\mathbf{w}_j, \mathbf{w}_i)) = 0.$$

Isso significa que  $T^*\mathbf{w}_j \perp \mathbf{w}_i$ , para todo  $i \neq j$ , de modo que  $T^*\mathbf{w}_j \in \text{span}\{\mathbf{w}_j\}$ . Isso significa que  $T^*\mathbf{w}_j$  é um múltiplo de  $\mathbf{w}_j$ , ou seja,  $\mathbf{w}_j$  é um autovetor de  $T^*$  e  $T^*\mathbf{w}_j = \mu_j\mathbf{w}_j$ , para algum autovalor  $\mu_j$ . Para descobirmos qual é esse autovalor, escolhemos  $i = j$  acima, de modo que

$$\mu_j = \mu_j\|\mathbf{w}_j\|^2 = ((\mu_j\mathbf{w}_j, \mathbf{w}_j)) = ((T^*\mathbf{w}_j, \mathbf{w}_j)) = \bar{\lambda}_j((\mathbf{w}_j, \mathbf{w}_j)) = \bar{\lambda}_j\|\mathbf{w}_j\|^2 = \bar{\lambda}_j.$$

Portanto,

$$T^*\mathbf{w}_j = \bar{\lambda}_j\mathbf{w}_j,$$

para todo  $j = 1, \dots, n$ . Nesse caso, expandindo um vetor  $\mathbf{u} \in X$  qualquer nessa base, i.e. escrevendo

$$\mathbf{u} = \sum_{j=1}^n z_j \mathbf{w}_j,$$

onde  $z_j \in \mathbb{C}$ , temos

$$\mathbf{T}^* \mathbf{T} \mathbf{u} = \sum_{j=1}^n z_j \mathbf{T}^* \mathbf{T} \mathbf{w}_j = \sum_{j=1}^n z_j \bar{\lambda}_j \lambda_j \mathbf{w}_j = \sum_{j=1}^n z_j \lambda_j \bar{\lambda}_j \mathbf{w}_j = \sum_{j=1}^n z_j \mathbf{T} \mathbf{T}^* \mathbf{w}_j = \mathbf{T} \mathbf{T}^* \mathbf{u},$$

provando que  $\mathbf{T}^* \mathbf{T} = \mathbf{T} \mathbf{T}^*$ , ou seja, que  $\mathbf{T}$  é normal.  $\square$

EXEMPLO 3.1. Em forma matricial, suponha que  $\mathbf{A} \in \mathbb{C}^{n \times n}$  seja uma matriz complexa com

$$\mathbf{U} \mathbf{A} \mathbf{U}^* = \mathbf{D},$$

onde  $\mathbf{U}$  é unitária (ou seja, suas colunas formam uma base ortonormal de  $\mathbb{C}^n$ ) e  $\mathbf{D}$  seja diagonal, com diagonal  $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ . Então

$$\mathbf{A} = \mathbf{U}^* \mathbf{D} \mathbf{U}$$

de modo que

$$\mathbf{A}^* = (\mathbf{U}^* \mathbf{D} \mathbf{U})^* = \mathbf{U}^* \mathbf{D}^* \mathbf{U}.$$

Como  $\mathbf{D}$  e  $\mathbf{D}^*$  são diagonais, então  $\mathbf{D}^* \mathbf{D} = \mathbf{D} \mathbf{D}^*$  (toda matriz diagonal é normal). Assim, usando que  $\mathbf{U} \mathbf{U}^* = \mathbf{I}$ ,

$$\mathbf{A}^* \mathbf{A} = \mathbf{U}^* \mathbf{D}^* \mathbf{U} \mathbf{U}^* \mathbf{D} \mathbf{U} = \mathbf{U}^* \mathbf{D}^* \mathbf{D} \mathbf{U} = \mathbf{U}^* \mathbf{D} \mathbf{D}^* \mathbf{U} = \mathbf{U}^* \mathbf{D} \mathbf{U} \mathbf{U}^* \mathbf{D}^* \mathbf{U} = \mathbf{A} \mathbf{A}^*.$$

Ou seja,  $\mathbf{A}$  é normal.

EXEMPLO 3.2. Considere a matriz complexa  $\mathbf{A} \in \mathbb{C}^2$  dada por

$$\begin{bmatrix} 2 + 3i & -1 \\ 1 & 2 + 3i \end{bmatrix}$$

Pode-se verificar que essa matriz é normal, i.e.  $\mathbf{A}^* \mathbf{A} = \mathbf{A} \mathbf{A}^*$ . Os autovalores são  $\lambda_1 = 2 + 4i$  e  $\lambda_2 = 2 + 2i$ . Uma base de autovetores é  $\mathcal{B} = \{\mathbf{w}_1, \mathbf{w}_2\}$ , onde

$$\mathbf{w}_1 = \frac{\sqrt{2}}{2}(1, -i), \quad \mathbf{w}_2 = \frac{\sqrt{2}}{2}(1, i).$$

Temos

$$\mathbf{A} \mathbf{w}_1 = \lambda_1 \mathbf{w}_1, \quad \mathbf{A} \mathbf{w}_2 = \lambda_2 \mathbf{w}_2.$$

Essa também é uma base de autovetores de  $\mathbf{A}^*$ , com

$$\mathbf{A} \mathbf{w}_1 = \bar{\lambda}_1 \mathbf{w}_1, \quad \mathbf{A} \mathbf{w}_2 = \bar{\lambda}_2 \mathbf{w}_2.$$

OBSERVAÇÃO 3.1. Conforme mencionado na demonstração do [Teorema 3.1](#), nem sempre um autovetor de  $T$  é autovetor de  $T^*$ . Isso acontece quando  $T$  é normal, mas nem sempre. E não é nada de patológico em relação a espaços complexos. Por exemplo, considere o operador associado à matriz real

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

cujo único autovalor é  $\lambda = 1$ . O autoespaço associado é o eixo  $x$ . Em particular,  $\mathbf{u} = (1, 0)$  é autovetor. No entanto, a adjunta é a transposta

$$\mathbf{A}^* = \mathbf{A}^{\text{tr}} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

e para a qual  $\mathbf{u} = (1, 0)$  não é autovetor. O autoespaço da matriz adjunta é o eixo  $y$ .

**3.2. Todo operador normal é diagonalizável em alguma base ortonormal.** Apresentamos, aqui, duas demonstrações desse importante resultado. A primeira demonstração começa como na demonstração da decomposição de Schur e explora o fato de que o subespaço ortogonal a um autoespaço é invariante pelo operador. A segunda demonstração usa a própria decomposição de Schur. Uma outra demonstração será feita quando fizermos a decomposição espectral de operadores hermitianos, pois é baseada no fato das partes real e imaginária de um operador normal comutarem entre si e, portanto, possuem uma base comum que as diagonaliza, nos dando uma diagonalização do próprio operador.

**TEOREMA 3.2.** *Seja  $T \in \mathcal{L}(X)$  um operador linear em um espaço vetorial complexo  $X$  com produto interno. Suponha que  $T$  seja normal e que  $X$  seja de dimensão finita. Então  $X$  possui uma base ortonormal de autovetores de  $T$ .*

**DEMONSTRAÇÃO DO [TEOREMA 3.2](#) VIA INDUÇÃO NA DIMENSÃO.** Como o espaço vetorial  $X$  é complexo, sabemos do [Teorema 3.1](#) que  $T$  possui pelo menos um autovalor, que denotamos por  $\lambda_1$ . Seja  $V(T, \lambda_1)$  o autoespaço associado a esse autovalor. Naturalmente,  $V(T, \lambda_1)$  é invariante por  $T$ . Vamos mostrar que subespaço perpendicular  $V(T, \lambda_1)^\perp$  também é invariante por  $T$ .

Para isso, vamos provar primeiro que  $V(T, \lambda_1)$  é invariante pelo operador adjunto  $T^*$ . Seja  $\mathbf{u} \in V(T, \lambda_1)$ . Usando que  $T^*$  é normal, temos

$$TT^*\mathbf{u} = T^*T\mathbf{u} = \lambda_1 T^*\mathbf{u},$$

mostrando que  $T^*\mathbf{u}$  é autovetor associado ao autovalor  $\lambda_1$ , ou seja  $T^*\mathbf{u} \in V(T, \lambda_1)$ , de forma que  $V(T, \lambda_1)$  é invariante por  $T^*$ .

Seja, agora,  $\mathbf{v} \in V(\mathbb{T}, \lambda_1)^\perp$ . Precisamos mostrar que  $\mathbb{T}\mathbf{v} \perp \mathbf{u}$  para todo  $\mathbf{u} \in V(\mathbb{T}, \lambda_1)$ . Para tal  $\mathbf{u}$ , já sabemos que  $\mathbb{T}^*\mathbf{u} \in V(\mathbb{T}, \lambda_1)$ . Com isso, pela ortogonalidade entre  $\mathbf{v}$  e  $V(\mathbb{T}, \lambda_1)$ , obtemos

$$((\mathbb{T}\mathbf{v}, \mathbf{u})) = ((\mathbf{v}, \mathbb{T}^*\mathbf{u})) = 0,$$

mostrando que  $\mathbb{T}\mathbf{v}$  também é ortogonal a  $V(\mathbb{T}, \lambda_1)$ .

Com isso, podemos restringir o operador  $\mathbb{T}$  ao subespaço perpendicular  $V(\mathbb{T}, \lambda_1)^\perp$ , reduzindo a dimensão e usando a hipótese de indução. De fato, seja  $S_1 = V(\mathbb{T}, \lambda_1)^\perp$  e seja  $\mathbb{T}_1 \in \mathcal{L}(S_1)$  a restrição de  $\mathbb{T}$  ao subespaço  $S_1$ . Como  $S_1$  é não trivial, a dimensão de  $S_1$  é estritamente menor do que a de  $X$ . Além disso,  $\mathbb{T}_1$  é normal. De fato, para  $\mathbf{v}, \mathbf{w} \in S_1$ , temos, por definição de adjunta,

$$((\mathbb{T}_1\mathbf{w}, \mathbf{v})) = ((\mathbf{w}, \mathbb{T}_1^*\mathbf{v})).$$

Por outro lado, como  $\mathbb{T}_1$  é a restrição de  $\mathbb{T}$  a  $S_1$ ,

$$((\mathbb{T}_1\mathbf{w}, \mathbf{v})) = ((\mathbb{T}\mathbf{w}, \mathbf{v})) = ((\mathbf{w}, \mathbb{T}^*\mathbf{v})).$$

Logo,

$$((\mathbf{w}, \mathbb{T}_1^*\mathbf{v} - \mathbb{T}^*\mathbf{v})) = 0.$$

É importante ressaltar que  $S_1$  é invariante por  $\mathbb{T}^*$ , também. De fato, se  $\mathbf{v} \in S_1$  e  $\mathbf{u} \in S_1$ , então

$$((\mathbf{u}, \mathbb{T}^*\mathbf{v})) = ((\mathbb{T}\mathbf{u}, \mathbf{v})) = 0,$$

visto que  $\mathbb{T}\mathbf{u} \in S_1$  e  $\mathbf{v} \in S_1$ . Assim, escolhendo  $\mathbf{w} = \mathbb{T}_1^*\mathbf{v} - \mathbb{T}^*\mathbf{v} \in S_1$  acima, obtemos

$$\|\mathbb{T}_1^*\mathbf{v} - \mathbb{T}^*\mathbf{v}\|^2 = ((\mathbb{T}_1^*\mathbf{v} - \mathbb{T}^*\mathbf{v}, \mathbb{T}_1^*\mathbf{v} - \mathbb{T}^*\mathbf{v})) = 0.$$

Logo,  $\mathbb{T}_1^* = \mathbb{T}^*|_{S_1}$ . Com isso,

$$\mathbb{T}_1^*\mathbb{T}_1 = \mathbb{T}^*|_{S_1} \mathbb{T}|_{S_1}.$$

Como  $S_1$  é invariante por ambos os operadores, temos

$$\mathbb{T}_1^*\mathbb{T}_1 = \mathbb{T}^*\mathbb{T}|_{S_1}.$$

Usando a normalidade de  $\mathbb{T}$ , temos

$$\mathbb{T}_1^*\mathbb{T}_1 = \mathbb{T}^*\mathbb{T}|_{S_1} = \mathbb{T}\mathbb{T}^*|_{S_1}.$$

Usando novamente a invariância,

$$\mathbb{T}_1^*\mathbb{T}_1 = \mathbb{T}|_{S_1} \mathbb{T}^*|_{S_1} = \mathbb{T}_1\mathbb{T}_1^*,$$

provando que  $T_1$  é normal. Pela hipótese de indução,  $S_1^\perp$  possui uma base ortonormal

$$\{\mathbf{w}_{n_1+1}, \dots, \mathbf{w}_n\},$$

de autovetores de  $T_1$ , i.e.

$$T_1 \mathbf{w}_j = \lambda_j \mathbf{w}_j,$$

para  $j = n_1+1, \dots, n$ , onde  $n_1 = \dim(S_1)$ , de modo que  $n - n_1 = \dim(S_1^\perp)$ . Escolhendo uma base ortonormal qualquer

$$\{\mathbf{w}_1, \dots, \mathbf{w}_{n_1}\}$$

de  $S_1$ , obtemos, visto que  $S_1$  e  $S_1^\perp$  são ortogonais, uma base ortonormal

$$\mathcal{B} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$$

de  $X$ . Nesse base, temos

$$T \mathbf{w}_j = \lambda_j \mathbf{w}_j,$$

para  $j = 1, \dots, n_1$  e, visto que  $T_1$  é a restrição de  $T$  a  $S_1^\perp$ ,

$$T \mathbf{w}_j = T_1 \mathbf{w}_j = \lambda_j \mathbf{w}_j,$$

para  $j = n_1 + 1, \dots, n$ . Isso completa a demonstração de existência de uma base ortonormal de autovetores de  $T$ .  $\square$

**OBSERVAÇÃO 3.2.** A demonstração via redução na dimensão através da invariância do ortogonal de um autoespaço está feita acima de forma bastante detalhada. Mas a ideia é simples. Passa por usar que (i) todo operador em um espaço vetorial complexo possui pelo menos um autovalor; (ii) o ortogonal de um autoespaço é invariante pelo operador; (iii) a restrição de um operador normal a um subespaço invariante é também normal.

Uma outra demonstração pode ser feita através da decomposição de Schur.

**DEMONSTRAÇÃO DO TEOREMA 3.2 VIA DECOMPOSIÇÃO DE SCHUR MATRICIAL.** Seja  $\mathbf{A} \in \mathbb{C}^{n \times n}$  uma matriz normal (representando  $T$  em uma base qualquer). Pelo **Teorema 2.2** da decomposição de Schur em forma matricial, existe uma matriz unitária  $\mathbf{U}$  e uma matriz triangular superior  $\mathbf{R}$  tais que

$$\mathbf{R} = \mathbf{U}^* \mathbf{A} \mathbf{U}.$$

Observe que  $\mathbf{R}$  também é normal, visto que

$$\mathbf{R} \mathbf{R}^* = \mathbf{U}^* \mathbf{A} \mathbf{U} \mathbf{U}^* \mathbf{A}^* \mathbf{U} = \mathbf{U}^* \mathbf{A} \mathbf{A}^* \mathbf{U} = \mathbf{U}^* \mathbf{A}^* \mathbf{A} \mathbf{U} = \mathbf{U}^* \mathbf{A}^* \mathbf{U} \mathbf{U}^* \mathbf{A} \mathbf{U} = \mathbf{R}^* \mathbf{R}.$$

Como  $\mathbf{R}$  é triangular superior, podemos escrever

$$\mathbf{R} = \begin{bmatrix} a_{11} & \mathbf{a}_{1,2:n} \\ \mathbf{0} & \mathbf{A}_{2:n,2:n} \end{bmatrix}$$

onde  $\mathbf{a}_{1,2:n} = (a_{12} \cdots a_{1n})$  e  $\mathbf{A}_{2:n,2:n} = (a_{ij})_{i,j=2}^n$ . Observe que

$$\mathbf{R}^* = \begin{bmatrix} \overline{a_{11}} & \mathbf{0}^* \\ \mathbf{a}_{1,2:n}^* & \mathbf{A}_{2:n,2:n}^* \end{bmatrix}$$

onde os asteriscos indicam as conjugadas transpostas dos vetores e matrizes em questão. Com isso, olhando apenas o primeiro elemento dos produtos das matrizes, temos

$$\mathbf{R}\mathbf{R}^* = \begin{bmatrix} |a_{11}|^2 + \|\mathbf{a}_{1,2:n}\|^2 & \star \\ \star & \star \end{bmatrix},$$

enquanto que

$$\mathbf{R}^*\mathbf{R} = \begin{bmatrix} |a_{11}|^2 & \star \\ \star & \star \end{bmatrix},$$

de modo que

$$\|\mathbf{a}_{1,2:n}\|^2 = 0,$$

ou seja  $\mathbf{a}_{1,2:n} = \mathbf{0}$  e  $\mathbf{R}$  tem a forma diagonal em blocos

$$\mathbf{R} = \begin{bmatrix} a_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{2:n,2:n} \end{bmatrix}$$

Prosseguindo por indução, chegamos a  $\mathbf{R}$  na forma diagonal

$$\mathbf{R} = \mathbf{D} = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \ddots & \vdots \\ \cdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_{nn} \end{bmatrix}.$$

Ou seja,  $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^*$ , mostrando que  $\mathbf{A}$ , e portanto  $\mathbb{T}$ , é diagonalizável, com base ortonormal de autovetores.  $\square$

**3.3. Equivalência.** Juntando os dois resultados anteriores, do [Teorema 3.1](#) e do [Teorema 3.2](#), obtemos a equivalência.

**COROLÁRIO 3.1.** *Seja  $X$  um espaço vetorial complexo com produto interno e de dimensão finita. Seja  $\mathbb{T} \in \mathcal{L}(X)$  um operador linear em  $X$ . Então  $\mathbb{T}$  é normal se, e somente se,  $X$  possui uma base ortonormal de autovetores de  $\mathbb{T}$ .*

Também temos a versão matricial.

**COROLÁRIO 3.2.** *Seja  $\mathbf{A} \in \mathbb{C}^{n \times n}$  uma matriz complexa, com  $n \in \mathbb{N}$ . Então  $\mathbf{A}$  é normal se, e somente se, existe uma matriz unitária  $\mathbf{U}$  e uma matriz diagonal  $\mathbf{D}$ , também em  $\mathbb{C}^{n \times n}$ , tais que*

$$\mathbf{A} = \mathbf{U}^*\mathbf{D}\mathbf{U}.$$

4. Diagonalização de operadores hermitianos

5. Operadores anti-hermitianos

6. Aplicações

6.1. Fórmula de Cauchy, resíduos e outras ferramentas de análise complexa em teoria espectral.

6.2. Séries de Fourier nos complexos.

6.3. Modelos em mecânica quântica. Ver [12, Chapter 10].

7. Exercícios

**Exercícios**

7.1. . Seja  $X$  um espaço vetorial complexo e suponha que  $b : X \times X \rightarrow \mathbb{C}$  seja uma forma bilinear e simétrica em  $X$ . Mostre que  $b$  não pode ser positiva definida.

## Referências Bibliográficas

- [1] S. Axler, *Linear Algebra Done Right*, 3rd edition, Springer, 2015.
- [2] G. Barenblatt, *Scaling*, Cambridge University Press, 2003.
- [3] W. M. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry*, Pure and Applied Mathematics, Volume 120, 2nd Edition, Academic Press, 2002.
- [4] N. Dunford & J. T. Schwartz, *Linear Operators I. General Theory*, With the assistance of W. G. Bade and R. G. Bartle, Pure and Applied Mathematics, Vol. 7, Interscience Publishers, Inc., New York, 1958.
- [5] M. B. Giles, (2008). Collected Matrix Derivative Results for Forward and Reverse Mode Algorithmic Differentiation. In: Bischof, C.H., Bücker, H.M., Hovland, P., Naumann, U., Utke, J. (eds) *Advances in Automatic Differentiation. Lecture Notes in Computational Science and Engineering*, vol 64. Springer, Berlin, Heidelberg, 2008. DOI: [10.1007/978-3-540-68942-3\\_4](https://doi.org/10.1007/978-3-540-68942-3_4)
- [6] E. van Groesen, J. Molenaar, *Continuum Modeling in the Physical Sciences*, SIAM, 2007.
- [7] S. Lang, *Linear Algebra*, 3rd edition, Springer, 1987.
- [8] P. Lax, *Linear Algebra and its Applications*, 2nd edition, Wiley-Interscience, 2007.
- [9] E. L. Lima, *Álgebra Linear*, IMPA, 2014.
- [10] Y. Ma, S. Gowda, R. Anantharaman, C. Laughman, V. Shah, C. Rackauckas, ModelingToolkit: A Composable Graph Transformation System For Equation-Based Modeling, [arXiv:2103.05244](https://arxiv.org/abs/2103.05244) (2021).
- [11] P. J. Olver; C. Shakiban, *Applied Linear Algebra*, 2nd edition, Springer, 2018.
- [12] I. Savov, *No Bullshit Guide to Linear Algebra*, 2nd edition, Minireference Co., 2020.
- [13] H. Shapiro, *Linear Algebra and Matrices - Topics for a Second Course*, American Mathematical Society, 2015.
- [14] T. Tao, A mathematical formalisation of dimensional analysis, [publicação no blog](#) *What's new*, 29 de dezembro de 2012.
- [15] A. Torchinsky, *Real Variables*, CRC Press, 1998.



## Símbolos e Abreviações

$[\cdot]$ : Classe de equivalência. 49

$\{\{\cdot\}\}$ : Símbolo para conjuntos ordenados e bases, e.g.  $\mathcal{b} = \{\{\mathbf{u}_1, \dots, \mathbf{u}_n\}\}$ . 37

$\mathcal{a}, \mathcal{b}, \mathcal{c}$ : Símbolos para representar conjuntos ordenados de vetores e bases. 37

$\mathbf{e}$ : Símbolo para representar base canônica. 37

$\delta_{ij}$ : Delta de Kronecker, definido por  $\delta_{ij} = 0$ , se  $i \neq j$ , e  $\delta_{ij} = \delta_{ii} = 1$ , quando  $i = j$ .  
22

$U, V, W, S, \dots$ : Símbolos para subespaços vetoriais. 27

$X/S$ : Espaço quociente, módulo subespaço  $S$ . 50

$X, Y, Z, \dots$ : Símbolos para espaços vetoriais. 15

$G(\mathbb{T}, \lambda)$ : Autoespaço generalizado de  $\mathbb{T}$  associado a um autovalor  $\lambda$ . 236

$V(\mathbb{T}, \lambda)$ : Autoespaço de  $\mathbb{T}$  associado a um autovalor  $\lambda$ . 218



## Índice Remissivo

Espaço vetorial, 14  
  espaço trivial, 17