Alguns resultados de Álgebra Linear

Caio Tomás de Paula
Matheus de Freitas Souza

Orientação: Prof. Dr. Lucas Conque Seco Ferreira

Departamento de Matemática

Universidade de Brasília

Publicações PETMAT

Outubro 2022

Resumo

Estas notas têm por objetivo registrar os tópicos discutidos durante a pesquisa individual do PETMAT realizada, pelos autores, ao longo do segundo semestre letivo de 2020. Os tópicos estudados vão desde uma definição mais adequada de subgrupo normal e uma discussão sobre quocientes, passando pelos subgrupos finitos de SO(3) e sua relação com os poliedros de Platão, pela simplicidade de SO(3), pela história e propriedades dos quatérnios e por uma demonstração alternativa do Teorema de Cayley-Hamilton utilizando o Teorema Fundamental dos Polinômios Simétricos.

Conteúdo

1 Homomorfismos, quocientes e normalidade

Para conhecer as estruturas de um grupo e subgrupos interessantes, podemos considerar funções entre grupos que preservam as operações: os homomorfismos. Sejam G e G' grupos, $\phi : G \to G'$ é um homomorfismo de grupos se, para todos $g, \tilde{g} \in G$, temos $\phi(g\tilde{g}) = \phi(g)\phi(\tilde{g})$. $H = \ker\phi$ é chamado de núcleo de ϕ e é composto dos $g \in G$ tais que $\phi(g) = e'$ (identidade de G'). A priori, ϕ não é necessariamente uma função bijetiva; considerando uma restrição $\psi : G \to \phi(G)$, definida por $\psi(g) = \phi(g)$ para todo $g \in G$, então ψ é homomorfismo sobrejetivo, por conta de sua definição e das propriedades que herda de ϕ.

Se ϕ não é injetiva, há ainda uma certa "redundância" nesta aplicação, no sentido de que há pontos distintos que têm a mesma imagem por ψ. Vamos ver que o núcleo $\ker\psi$ do homomorfismo é uma ferramenta importante na tentativa de eliminar esta "redundância".

Note que se $g, \tilde{g} \in G$ e $\psi(g) = \psi(\tilde{g})$, então

$$e' = \psi(g^{-1}\tilde{g}).$$

Esta relação nos leva a $g^{-1}\tilde{g} \in H$. Por este motivo, definimos a seguinte relação de equivalência em G:

$$gR\tilde{g} \iff \psi(g) = \psi(\tilde{g}) \iff g^{-1}\tilde{g} \in H.$$

Dado $g \in G$, defina $A_g = \{x \in G : \psi(x) = \psi(g)\}$ e considere as classes laterais gH e Hg de H. Vamos mostrar que estes conjuntos são todos iguais. É evidente que $gH \subseteq A_g$ e $Hg \subseteq A_g$. Tomando $x \in A_g$ qualquer, temos $xg^{-1} \in H$ e $g^{-1}x \in H$, o que mostra que $A_g \subseteq gH$ e $A_g \subseteq Hg$. Logo, $gH = Hg$. Esta igualdade nos permite concluir ainda que $gHg^{-1} = H$ e a volta também é válida.

Estamos separando os pontos com mesma imagem nas classes laterais $gH(= Hg)$, que determinam uma partição do conjunto G. Denotemos por $G/H = \{gH : g \in G\}$ o conjunto das partições. Defina a aplicação $\pi : G \to G/H$, dada por $\pi(g) = gH$. Vamos ver que esta função é um homomorfismo sobrejetivo.

O fato de que $H = \ker\psi$ nos permite dar uma estrutura de grupo a G/H definindo uma operação entre seus elementos. Defina

$$(aH)(bH) = (ab)H. \tag{*}$$

Para que esta seja de fato uma operação razoável para um grupo, precisamos que ela esteja bem definida. Vamos mostrar que isso acontece se, e somente se, $gH = Hg$.

Se vale a equação em (*) e $gHg^{-1} = H$, sejam $a' \in aH, b' \in bH$; neste caso, $a'H = aH$ e $b'H = bH$. Então $a' = ah_1$ e $b' = bh_2$, para certos $h_1, h_2 \in H$. Devemos mostrar que $(a'H)(b'H) = (aH)(bH)$ (boa definição). Observe que

$$(a'H)(b'H) = (a'b')H = (ah_1bh_2)H = (ah_1b)H = (abh_0)H = (ab)H,$$

onde $bh_0 = h_1b$ para algum $h_0 \in G$, uma vez que $bH = Hb$. Portanto, esta operação está bem definida. Reciprocamente, suponha agora que a operação está bem definida. Neste caso, temos que

$$(eH)(gH) = gH,$$

ou seja, existem $h_0, h_1, h_2 \in H$ tais que $h_1 g h_2 = g h_0$, i.e., $h_1 g = g h_0 h_2^{-1}$. Donde segue que $gH = Hg$, ou, equivalentemente, $gHg^{-1} = H$.

Com esta operação definida, G/H é um grupo. Podemos definir então um isomorfismo $\gamma : G/H \to \phi(G)$, dado por $\gamma(gH) = \psi(g)$. Este é de fato um isomorfismo por conta da propriedade de homomorfismo e da sobrejetividade de ψ, além da forma como G/H foi construído e do modo que sua operação foi definida. Um subgrupo H de um grupo G com estas propriedades é chamado **subgrupo normal** de G, denotado $H \triangleleft G$.

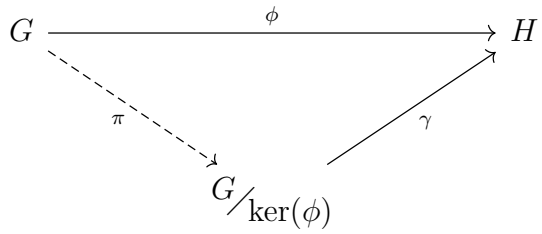


Figura 1: Diagrama de um subgrupo normal.

Daí, a seguinte definição segue naturalmente:

Definição 1.1. *Um subgrupo de um grupo G é chamado normal se é núcleo de um homomorfismo.*

Segue desta definição que um subgrupo é normal se, e somente se, $gHg^{-1} = H$. Observe ainda que G e $\{e\}$ são sempre subgrupos normais de G.

Vamos agora tratar de uma generalização deste resultado. Consideramos um grupo G agindo sobre um conjunto X. Podemos enxergar a ação de G em X como uma aplicação $\Psi : G \times X \to X$ que, a cada par (g, x), associa um ponto $\Psi(g, x) \in X$ que denota a ação de g em x. Para uma notação mais concisa, $\Psi(g, x) = gx$. Um grupo G como esse é dito ter uma ação transitiva se, intuitivamente, sua ação leva qualquer ponto de X em qualquer outro ponto que se queira. Mais precisamente, dado $x \in X$, defina $Gx = \{gx : g \in G\}$, chamada **órbita** do ponto x. Então a ação de G é transitiva se $Gx = X$.

O subconjunto de G que deixa invariante um dado elemento $x \in X$ é chamado **estabilizador** de x ou **isotropia** de G em x e denotado por $\mathrm{Stab}(x) = \{g \in G : gx = x\}$. Veremos que este conjunto tem um papel similar ao do conjunto $H = \ker \phi$ que tratamos anteriormente.

Para uma discussão mais interessante, suponhamos que G não tenha ação transitiva. Sabemos que Ψ é uma aplicação sobrejetiva, já que G é um grupo e possui um elemento identidade e e, pela definição de ação de grupo, temos $\Psi(e, x) = x$, $\forall\ x \in X$. Se Ψ é injetiva, então $\mathrm{Stab}(x)$ tem, no máximo, dois elementos, caso haja algum elemento distinto da identidade que fixe x. Vamos supor, de modo geral, que esta aplicação não seja injetiva.

Neste caso, há novamente uma certa "redundância" na imagem dos pontos de X, i.e., há elementos $g, g' \in G$ tais que $gx = g'x$. Para eliminar este tipo de situação, o estabilizador é essencial. Considerando a órbita Gx de x, pode-se checar que o conjunto das órbitas dos elementos de X determina uma partição de X, e isto nos possibilita

trabalhar em um conjunto reduzido de X. Não obstante, os resultados que obteremos valem para todos os pontos de X.

Note que $gx = g'x$ implica $x = g^{-1}g'x$, ou seja, $g^{-1}g' \in \mathrm{Stab}(x)$. Novamente, podemos relacionar todos os pontos que levam x em um mesmo ponto gx. Defina

$$gRg' \iff gx = g'x \iff g^{-1}g' \in \mathrm{Stab}(x).$$

Considere o conjunto das partições determinadas por esta relação, isto é, o conjunto

$$G/\mathrm{Stab}(x) = \{g\,\mathrm{Stab}(x) : g \in G\}.$$

Este conjunto não é necessariamente um grupo, pois apesar de ser subgrupo, $\mathrm{Stab}(x)$ não é necessariamente normal em G, i.e., não se pode garantir que ele seja o núcleo de um homomorfismo partindo de G.

Definimos então uma aplicação $\pi : G \to G/\mathrm{Stab}(x)$ dada por $\pi(g) = g\,\mathrm{Stab}(x)$ (não é um homomorfismo!). Com esta construção, podemos definir a aplicação bijetiva $\Gamma : G/\mathrm{Stab}(x) \times X \to X$, dada por $\Gamma(gH, x) = \Psi(g, x) = gx$, que age sobre os elementos de X.

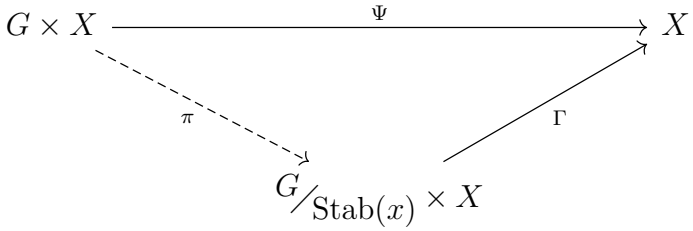


Figura 2: Diagrama da ação de Ψ e Γ.

Além disso, há certas conclusões bastante interessantes e úteis. Por exemplo, cada elemento de $G/\mathrm{Stab}(x)$ está relacionado de maneira biunívoca com um ponto da órbita Gx, ou seja, cada classe lateral $g\,\mathrm{Stab}(x)$ age da mesma maneira sobre x. Segue que

$$\begin{aligned} |G| &= \sum_{G/\mathrm{Stab}(x)} |g\,\mathrm{Stab}(x)| \\ &= |Gx||\,\mathrm{Stab}(x)|, \end{aligned}$$

identidade que nos livros-texto é chamada de **Teorema Órbita-Estabilizador**.

2 Subgrupos finitos de SO(3) e os poliedros de platão

Começamos esta seção descrevendo alguns subgrupos finitos de SO(3) e algumas figuras sobre as quais eles agem. Para o leitor não familiarizado, o grupo SO(3) nada mais é do que o grupo de rotações em torno da origem no $\mathbb{R}^3$. O grupo cíclico de ordem n, i.e., C_n, é o grupo de rotações em torno de um mesmo eixo, e seu gerador é a rotação de menor ângulo $2\pi/n$. Este grupo é o grupo de simetrias de objetos do seguinte tipo, que é muitas vezes denominado *grupo monoedral.*

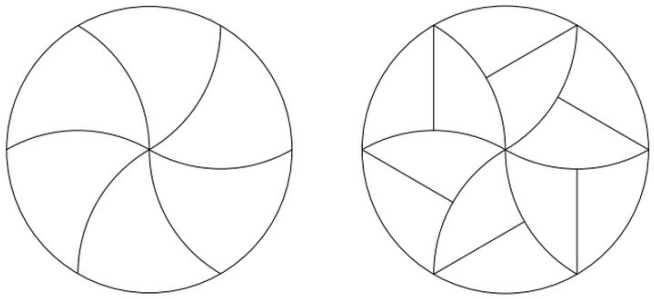


Figura 3: Exemplos de figuras sobre as quais o grupo monoedral age.

Um outro subgrupo importante e mais conhecido é o *grupo diedral*, D_n, que é o grupo de simetrias dos polígonos regulares de n lados. Diferentemente do monoedral, este contém rotações e reflexões por retas que passam por um vértice e pelo centro do polígono regular. Além disso, este grupo é gerado pela rotação de menor ângulo $2\pi/n$ e por um reflexão qualquer do grupo. Aqui, as reflexões se expressam como rotações de π pelo eixo que nos referimos como de "reflexão".

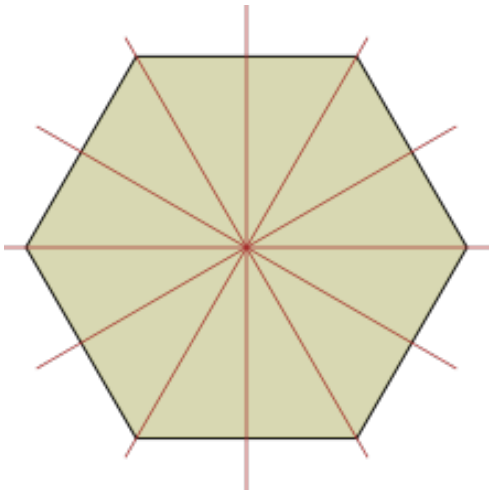


Figura 4: Hexágono regular com seus eixos de simetria.

Determinamos outros grupos finitos de SO(3) considerando poliedros regulares conhecidos e suas simetrias. Os **sólidos de Platão** são sólidos que têm, como faces, polígonos regulares. Há 5 desses poliedros: tetraedro, cubo, octaedro, icosaedro e dodecaedro.

O tetraedro é uma pirâmide triangular em que todas as faces são triângulos equiláteros, como ilustrado abaixo.

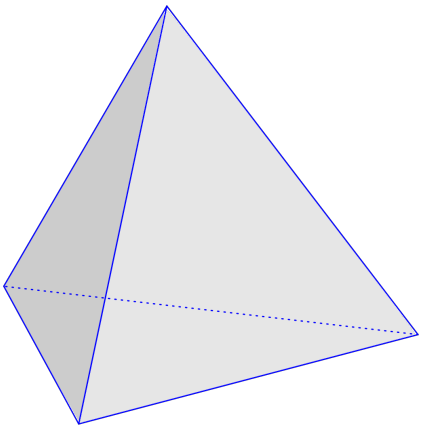


Figura 5: Tetraedro regular.

O grupo de simetrias que preservam orientação do tetraedro consiste de rotações de π por eixos que passam por pontos médios de vértices opostos e de rotações de $2\pi/3$ por eixos que que passam por um vértice e pelo centro da face oposta a ele. Este é um grupo de ordem 12, que denotaremos por T.

O cubo é um paralelepípedo que tem todas as faces quadradas. O octaedro é construído tomando-se como seus vértices os pontos centrais das faces do cubo e pode ser visto como duas pirâmides quadradas com bases coincidentes.

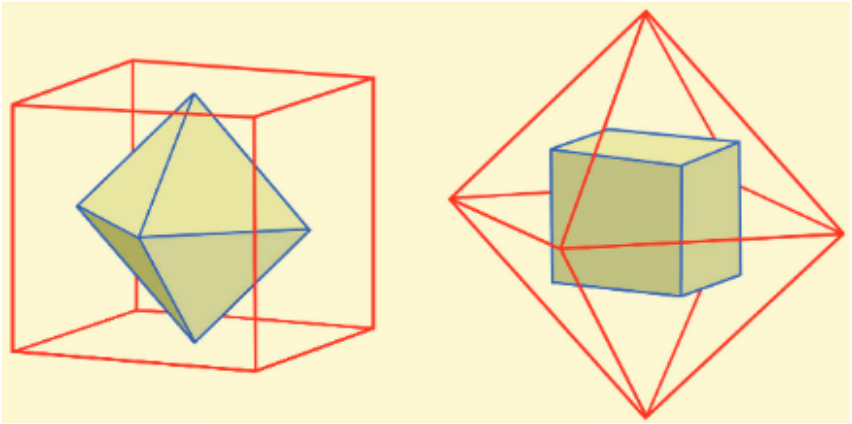


Figura 6: A dualidade entre cubo e octaedro.

O grupo de simetrias do cubo consta de rotações de $\pi/2, \pi$ e $2\pi/3$ por eixos que passam por centros de faces opostas, pontos médios de arestas opostas e por vértices opostos, respectivamente. Este grupo tem ordem 24, e será denotado por O. O mesmo grupo, aqui descrito para o cubo, é o grupo de simetrias positivas do octaedro por sua construção.

O icosaedro é um poliedro com 20 faces que são triângulos equiláteros. O dodecaedro é um poliedro com 12 faces que são pentágonos regulares e pode ser construído tomando como vértices os centros das faces do icosaedro.

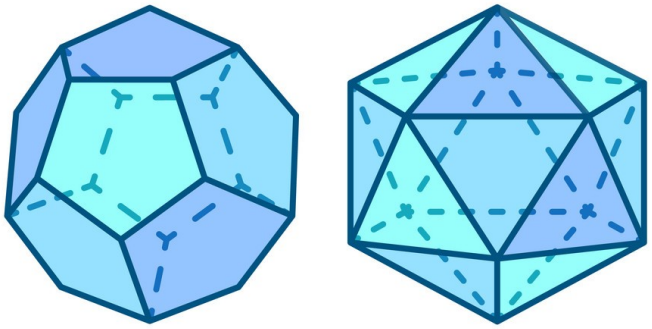


Figura 7: O dodecaedro (esquerda) e o icosaedro (direita).

Considere o dodecaedro. O seu grupo de simetrias positivas, aqui denotado por I, consta de rotações de $2\pi/5, 2\pi/3$ e π em torno de eixos por centros de faces opostas, por vértices opostos e por pontos médios de vértices opostos, respectivamente.

Agora, usando o resultado obtido anteriormente que

$$|G| = |Gx||\operatorname{Stab}(x)|,$$

vamos mostrar que há apenas 5 subgrupos finitos de SO(3) a menos de isomorfismos. Como corolário vamos concluir que há apenas 5 poliedros regulares (chamados sólidos de Platão). Estes subgrupos são exatamente os grupos de simetrias dos polígonos regulares que preservam orientação, i.e., são grupos de rotação. Os grupos e os objetos em que agem são listados abaixo:

- C_n: Grupo cíclico de ordem n.
- D_n: Grupo diedral n.
- T: Grupo de simetrias de um tetraedro regular.
- O: Grupo de simetrias do cubo e do octaedro.
- I: grupo de simetrias do icosaedro e do dodecaedro.

Teorema 2.1. *Os subgrupos finitos de* SO(3) *são apenas* C_n, D_n, T, O *e* I.

Demonstração. Os elementos de SO(3) são todos rotações em torno de algum eixo pela origem de $\mathbb{R}^3$. Podemos considerar, para facilitar a discussão, a esfera unitária

$$S^2 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$$

e pensar em como as rotações agem sobre ela.

É fácil ver que SO(3) é um grupo de simetrias da esfera e que cada rotação distinta da identidade fixa exatamente dois pontos: os pontos de interseção da esfera com o eixo de rotação. Estes pontos são chamados polos e serão denotados por p. Seja então P o conjunto de todos os pontos p tais que p é polo de algum elemento de G, onde G é um subgrupo qualquer de SO(3) finito com ordem N. O seguinte lema é importante para o que vem a seguir.

Lema 2.2. *O conjunto de polos é levado nele mesmo, ou seja, os elementos de G são operadores em P.*

Demonstração. Se $g \in G$ é um elemento qualquer e p é um polo de g, devemos mostrar que existe $x \in G$ tal que xp é um polo, i.e., que existe algum elemento de G que fixa xp. Este elemento é exatamente $xgx^{-1} \in G$. □

Este resultado estabelece que G deve ser um grupo de simetrias do conjunto de polos de seus elementos. Denote $r_p = |\operatorname{Stab}(p)|$ e $n_p = |Gp|$. Temos então

$$N = r_p n_p \iff |G| = |\operatorname{Stab}(p)||Gp|.$$

O argumento que segue se baseia na contagem da quantidade de polos que um subgrupo do tipo de G pode ter e na ordem das órbitas relacionadas com esses polos.

Observe que r_p é a ordem do subgrupo cíclico gerado pela rotação de menor ângulo $2\pi/r_p$ e que tem polo em p. Então cada polo está relacionado com exatamente $r_p - 1$ elementos do grupo. Além disso, o polo antípoda a p está relacionado com esses mesmos elementos de G. Portanto, retirando o elemento identidade, temos

$$\sum_{p \in P} (r_p - 1) = 2(N - 1). \tag{$\star$}$$

Se p e p' estão na mesma órbita, então $Gp = Gp'$ e $n_p = n_{p'}$. Logo, $r_p = r_{p'}$ pois uma órbita determina uma partição do conjunto em que o grupo age. Podemos, então, realizar o somatório sobre as órbitas ao invés de realizá-la sobre o conjunto de polos. Procedendo desta forma, obtemos

$$\sum_{Gp} n_p(r_p - 1) = 2N - 2$$

e, dividindo por N, segue que

$$\sum_{Gp} \left(1 - \frac{1}{r_p}\right) = 2 - \frac{2}{N}. \tag{$\star\star$}$$

O fato chave a ser observado é que

$$1 \leq 2 - \frac{2}{N} < 2.$$

Como $\operatorname{Stab}(p)$ deve conter um elemento distinto da identidade, então $r_p \geq 2$. O lado esquerdo em ($\star\star$) tem termos que são pelo menos $1/2$, logo, temos no máximo 3 órbitas. Além disso, não podemos ter 1 órbita apenas, pois

$$1 - \frac{1}{r_p} < 1$$

não concorda com o lado direito de ($\star\star$). Temos, portanto, 2 casos a considerar, número de órbitas 2 ou 3. Para os argumentos que seguem, denote as órbitas por números $1, 2, 3$. Então, ao invés de denotar r_p para p em Gp, denotemos r_1 e assim por diante. Suponha, ainda, que tenhamos $r_i \geq r_{i+1}$.

I) **Duas órbitas.** Neste caso, temos

$$\frac{2}{N} = \frac{1}{r_1} + \frac{1}{r_2}.$$

Como r_1 e r_2 dividem N, então $r_1, r_2 \leq N$. Segue, portanto, que $r_1 = r_2 = N$. Observe que temos duas órbitas de ordem 1 (apenas um elemento em cada órbita!) e os estabilizadores dessas órbitas têm ordem N. Os pontos destas órbitas devem determinar um eixo de rotação que é o eixo de rotação dos elementos de um subgrupo cíclico de ordem N, ou seja, C_N.

II) **Três órbitas.**

Este é o caso principal, aqui temos mais possibilidades. Veja que se r_1, r_2, r_3 são pelo menos 3, então

$$2 > 2 - \frac{2}{N} = \sum_{i=1}^{3} \left(1 - \frac{1}{r_i}\right) = 3 - \sum_{i=1}^{3} \frac{1}{r_i} \geq 3 - 1 = 2,$$

o que é absurdo. Logo, um dos r_i é 2, digamos $r_1 = 2$. Os casos possíveis estão listados abaixo. Denote, a partir daqui, as órbitas por O_1, O_2 e O_3.

	r_1	r_2	r_3	$\lvert G\rvert$
(i)	n	n	-	n
(ii)	2	2	n	$2n$
(iii)	2	3	3	12
(iv)	2	3	4	24
(v)	2	3	5	60

(ii) Aqui temos $r_1 = 2 = r_2$ e $n_1 = n = n_2$, então temos 2 estabilizadores com apenas 2 elementos e órbitas de n pontos. Como $r_3 = n$ e $n_3 = 2$, temos uma órbita com 2 elementos. Os elementos de O_3 devem ser permutados, o que indica uma rotação de π radianos. Além disso, os estabilizadores com 2 elementos são exatamente os subgrupos que contêm esta rotação (que pode ser interpretada também como uma reflexão numa certa reta contida num plano!).

Os elementos das órbitas de n pontos determinam exatamente um polígono regular e estão relacionados com o estabilizador que contém n elementos. Observe que a órbita O_3, neste caso, determina o eixo de rotação do subgrupo de ordem n. Portanto, $G \simeq D_n$.

(iii) Lembre-se das descrições das simetrias do tetraedro regular. Temos um subgrupo cíclico de G com 2 elementos, é o que nos diz $r_1 = 2$. Este se comporta como a simetria de rotação com eixo nos pontos médios de vértices opostos do tetraedro.

Além disso, temos dois subgrupos cíclicos de 3 elementos e duas órbitas de 4 pontos. Observe que, no tetraedro, qualquer um dos 4 vértices podem ser levados uns nos outros compondo-se as rotações que são simetrias das faces. Lembrando das descrições das simetrias do tetraedro regular, concluímos que $G \simeq T$.

(iv) Temos $r_1 = 2$, o que indica um subgrupo de G de ordem 2 (rotação de π, um grupo que tem um elemento involutivo). Como $r_2 = 3$ e $n_2 = 8$, temos um subgrupo cíclico de ordem 3 e uma órbita de ordem 8. Analogamente, $r_3 = 4$ caracteriza uma subgrupo cíclico de ordem 4.

Os valores de r_1, r_2 e r_3 caracterizam, respectivamente, as simetrias de rotação com eixos por pontos médios de arestas opostas, por vértices opostos e por centros de faces opostas. Observe que a órbita O_2 de ordem 8 é exatamente o conjunto de vértices do cubo. Segue que $G \simeq O$.

(v) Aqui temos como fato mais marcante a existência de um subgrupo cíclico de ordem $5 = r_3$, que está relacionado com a simetria rotacional por um eixo que passa por centros de faces opostas.

Temos uma órbita com $20 = n_2$ elementos, exatamente o número de vértices de um dodecaedro. Além disso, um subgrupo de ordem $2 = r_1$ deve conter uma rotação de π; este está relacionado com rotações por pontos médios de vértices opostos. Por fim, temos um subgrupo cíclico de ordem 3 que é gerado por rotações de $2\pi/3$ ao redor de um eixo por vértices opostos, e segue que $G \simeq I$.

□

3 A simplicidade de SO(3)

Dizemos que um grupo G é **simples** se os seus únicos subgrupos normais são $\{1\}$ e G. A seguir, mostramos a simplicidade de SO(3).

Teorema 3.1. *O único subgrupo não trivial de* SO(3) *fechado sob conjugação é o próprio* SO(3).

Demonstração. Suponha que H seja um subgrupo não trivial de SO(3). Assim, H contém uma rotação não trivial, h, de um ângulo α em torno do eixo l.

Agora, suponha que H é normal, i.e., H contém todos os elementos da forma ghg^{-1}, com $g \in \mathrm{SO}(3)$. Se g leva l em m, então g^{-1} leva m em l e h fixa l e rotaciona o restante dos pontos por α, segue que ghg^{-1} é uma rotação de α em torno de m. Portanto, H inclui as rotações de α em torno de *todos os eixos possíveis*, já que m era arbitrário; para ver isto, basta notar que a ação de $SO(3)$ sobre os pontos de S^2 é transitiva, logo, l pode ser levado em qualquer eixo m dado que passa pela origem.

Escreva R_α^P para indicar uma rotação de α em torno do eixo que passa por P e pela origem (como na figura abaixo) e R_{PQ} para indicar a reflexão em torno do grande círculo PQ, que passa por P e Q. Daí, temos

$$R_\alpha^Q R_\alpha^P = R_\theta^R,$$

em que R e θ são como mostrado abaixo.

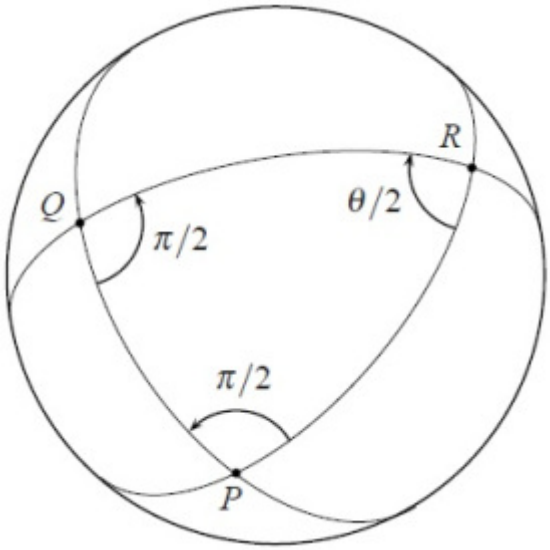


Figura 8: Ângulo da rotação produto.

Podemos, ainda, escrever

$$R_\theta^R = R_\alpha^Q R_\alpha^P = R_{QR}R_{PQ}R_{PQ}R_{PR} = R_{QR}R_{PR},$$

ou seja, a rotação em torno de R é definida pelas reflexões em torno de PR e QR.

À medida que P varia continuamente em algum intervalo do grande círculo por P e Q, o ângulo θ varia continuamente em algum intervalo (R pode variar também, mas isso não muda a análise). Como os números da forma

$$\frac{m\pi}{n}, m, n \in \mathbb{Z} \text{ e } m \text{ ímpar},$$

são densos em $\mathbb{R}$, segue que θ assume um valor dessa forma. Fazendo o produto dessa rotação consigo mesma n vezes obtemos uma rotação de $m\pi$ em torno de R, com m ímpar, ou seja, uma rotação de π em torno de R. Como H é fechado e R é arbitrário, segue que H contém todas as rotações de π em torno de qualquer ponto da esfera.

Por fim, tomando o produto de rotações da figura com $\alpha/2 = \pi/2$, podemos obter uma rotação em torno de R de qualquer $0 \leq \theta \leq 2\pi$: de fato, fixando Q e fazendo P variar sobre o círculo PQ na Figura 8, podemos ver que $\theta/2$ percorre todo o intervalo $[0, \pi]$. Assim, H contém todas as rotações de SO(3), i.e., $H = \text{SO}(3)$. □

4 Quatérnios e rotações no espaço

Vamos, nesta seção, introduzir o conjunto dos quatérnios, denotado $\mathbb{H}$ em homenagem ao matemático William Hamilton, que os descobriu. Focaremos sobretudo nos aspectos geométricos e propriedades básicas desse conjunto (munido das operações adequadas). Sendo assim, uma primeira motivação interessante para o estudo dos quatérnios é a que apresentaremos a seguir.

De modo relativamente informal, podemos considerar a seguinte definição.

Definição 4.1. *Dado um corpo K, um espaço vetorial V sobre K com produto bilinear é chamado álgebra real.*

Esse conceito pode parecer bastante fora da realidade, mas na tabela abaixo citamos alguns exemplos de álgebras com as quais a maioria de nós está acostumado.

	pol. reais	$(C(\mathbb{R}),\cdot)$	$(C_0(\mathbb{R}),*)$	$\mathbb{R}$	$\mathbb{C}$	$\mathbb{H}$	$\mathbb{O}$	$(\mathbb{R}^3,\times)$
Id.	✓	✓	✗	✓	✓	✓	✓	✗
Assoc.	✓	✓	✓	✓	✓	✓	✗	✗

Exemplos de álgebras reais.

Dizemos, ainda, que uma álgebra é **de divisão** se dados quaisquer dois elementos a e $b \neq 0$ nela, existem únicos x, y também na álgebra tais que $a = bx = yb$. Nesse contexto, temos o

Teorema 4.1 (Frobenius)**.** *Toda álgebra de divisão com dimensão finita e associativa sobre $\mathbb{R}$ é isomorfa a $\mathbb{R}, \mathbb{C}$ ou $\mathbb{H}$, de dimensões $1, 2$ e 4, respectivamente.*

Retirando a hipótese de associatividade, temos ainda uma quarta possibilidade: o conjunto $\mathbb{O}$ dos octônios. Nesse contexto, podemos ver que os quatérnios aparecem de maneira natural num teorema de classificação importante, despertando a suspeita de que talvez eles de fato sejam razoavelmente relevantes. De fato, o conjunto dos quatérnios é bastante útil para descrever algo que, ironicamente, é extremamente cotidiano: rotações no espaço tridimensional.*

Foi daí que o matemático Hamilton partiu para descobrir esse novo conjunto numérico. Ele buscava uma forma de descrever rotações em três dimensões similar à maneira que descrevemos rotações no plano por meio de números complexos, e acabou encontrando os quatérnios, definidos abaixo. Antes, porém, demonstramos um teorema que nos será útil ao final da seção.

Teorema 4.2 (Euler)**.** *Todo $A \in \mathrm{SO}(3)$ é uma rotação ao redor de um eixo.*

Demonstração. Note que é suficiente mostrar que A tem um autovalor 1 pois, se isso acontece, então A fixa uma direção, i.e., um eixo. O polinômio característico de uma matriz de $\mathrm{SO}(3)$ tem grau 3, então cada matriz desse grupo tem pelo menos uma raiz real. Usando o fato de que toda matriz ortogonal preserva produto interno e, consequentemente, preserva a norma, temos que, para toda matriz $A \in \mathrm{SO}(3)$, vale

$$|v| = |Av| = |\lambda v| = |\lambda||v|.$$

Então, sendo λ um autovalor de $A \in \mathrm{SO}(3)$, temos $|\lambda| = 1$. Como as transformações de $\mathrm{SO}(3)$ são aquelas que preservam orientação, necessariamente $\lambda = 1$; caso contrário $Av = -v$ e A não preservaria orientação. □

Definição 4.2. *O conjunto $\mathbb{H}$ dos quatérnios é definido por*

$$\mathbb{H} = \{a + bi + cj + dk : a, b, c, d \in \mathbb{R}\}$$

com a operação usual de soma. As constantes imaginárias i, j e k são tais que

$$i^2 = j^2 = k^2 = ijk = -1.$$

*O leitor interessado é convidado a assistir a este vídeo, que ilustra de maneira magnífica a vantagem de se trabalhar com quatérnios para descrever rotações tridimensionais.

Com essa definição, podemos ver que $\mathbb{H}$ é um espaço vetorial de dimensão 4. Além disso, existe uma cópia de $\mathbb{R}$ em $\mathbb{H}$. Também podemos ver que 1 é a identidade e o produto das unidades imaginárias é associativo (basta fazer as várias combinações). Então, antes de definir o produto quaterniônico, um esclarecimento de notação: os elementos do subconjunto

$$\mathrm{Im}(\mathbb{H}) = \{bi + cj + dk : b, c, d \in \mathbb{R}\}$$

de $\mathbb{H}$ são chamados **imaginários** e, dado um quatérnio

$$p = a + bi + cj + dk,$$

a tripla (b, c, d) é a **parte imaginária** de p.

Então, tome inicialmente $p, q \in \mathrm{Im}(\mathbb{H})$. Temos

$$\begin{aligned} p \cdot q &= (ai + bj + ck)(\alpha i + \beta j + \delta k) \\ &= -\langle p, q\rangle + i(b\delta - c\beta) + j(c\alpha - a\delta) + k(a\beta - b\alpha) \\ &= -\langle p, q\rangle + \begin{vmatrix} a & b & c \\ \alpha & \beta & \delta \\ i & j & k \end{vmatrix}, \end{aligned}$$

usando o produto canônico de $\mathbb{R}^3$. Assim, a seguinte definição é mais que natural.

Definição 4.3. *O produto quaterniônico $(\cdot)$ entre dois quatérnios imaginários p, q é definido por*

$$p \cdot q = p \times q - \langle p, q\rangle.$$

Aqui é interessante fazer algumas observações. Da fórmula acima, podemos ver que o produto quaterniônico entre imaginários é composto de duas partes: uma parte simétrica (o produto escalar) e uma parte anti-simétrica (o produto vetorial). Essa é a manifestação geométrica da anti-simetria do produto entre as unidades imaginárias i, j e k.

Já podemos deduzir algumas igualdades interessantes. Primeiro, note que se $p \in \mathrm{Im}(\mathbb{H})$, então

$$p \cdot p = p \times p - \langle p, p\rangle = -|p|^2$$

e, se p tem norma 1, segue que

$$p^2 = -1,$$

em analogia a $\mathbb{C}$.

Para terminar de definir o produto quaterniônico, precisamos lidar com quatérnios gerais. Então, tomando $p, q \in \mathbb{H}$ quaisquer, o produto entre eles é dado por

$$\begin{aligned} p \cdot q &= (\mathrm{Re}(p) + \mathrm{Im}(p))(\mathrm{Re}(q) + \mathrm{Im}(q)) \\ &= \mathrm{Re}(p)\mathrm{Re}(q) + \mathrm{Im}(p) \times \mathrm{Im}(q) - \langle \mathrm{Im}(p), \mathrm{Im}(q)\rangle + \mathrm{Re}(p) \times \mathrm{Im}(q) + \mathrm{Re}(q) \times \mathrm{Im}(p). \end{aligned}$$

Ainda se tratando do produto quaterniônico, podemos definir uma operação de colchete $[\cdot, \cdot]$ nos quatérnios por

$$[p, q] = p \cdot q - q \cdot p.$$

Assim, se $p, q \in \mathrm{Im}(\mathbb{H})$, então

$$[p, q] = p \times q - \langle p, q\rangle - q \times p + \langle q, p\rangle = 2(p \times q).$$

Claramente, essa operação de colchete não é simétrica. Veremos mais à frente que ela está intimamente relacionada com as rotações em $\mathbb{R}^3$.

Agora, definindo o conjugado quaterniônico da maneira natural, i.e.,

$$\overline{p} = \overline{a + bi + cj + dk} = a - bi - cj - dk,$$

é imediato que

$$\begin{aligned}\overline{i} &= -i,\\ \overline{j} &= -j,\\ \overline{k} &= -k,\end{aligned}$$

donde segue que

$$\overline{p} = a + b\overline{i} + c\overline{j} + d\overline{k},$$

de modo que a correspondência $p \mapsto \overline{p}$ é $\mathbb{R}$-linear. Além disso, note que

$$\overline{\overline{p}} = p,$$

analogamente a $\mathbb{C}$. Agora, diferentemente de $\mathbb{C}$, temos, por exemplo

$$\overline{ij} = \overline{k} = -k = ji = \overline{j} \cdot \overline{i},$$

de modo que $\overline{pq} \neq \overline{p} \cdot \overline{q}$ em geral. A igualdade acima sugere, na verdade, que $\overline{pq} = \overline{q} \cdot \overline{p}$. De fato, isso segue da igualdade acima e das igualdades

$$\begin{aligned}\overline{jk} = \overline{i} = -i = kj = \overline{k} \cdot \overline{j},\\ \overline{ki} = \overline{j} = -j = ik = \overline{i} \cdot \overline{k}.\end{aligned}$$

Resta agora verificar as propriedades do valor absoluto para quatérnios, que é definido da maneira natural:

$$|p| = \sqrt{a^2 + b^2 + c^2 + d^2}.$$

Assim como em $\mathbb{C}$, teremos $|p|^2 = p\overline{p}$. De fato, isso pode ser verificado usando a fórmula geral para o produto e notando que $\mathrm{Im}(\overline{p}) = -\mathrm{Im}(p)$, ou notando que

$$\overline{p\overline{p}} = \overline{\overline{p}} \cdot \overline{p} = p\overline{p},$$

ou seja, $p\overline{p} \in \mathrm{Re}(\mathbb{H})$. Como $\mathrm{Re}(p\overline{p}) = |p|^2$, segue que $p\overline{p} = |p|^2$. Uma terceira forma de ver isso é notar que

$$\begin{aligned}p\overline{p} &= (\mathrm{Re}(p) + \mathrm{Im}(p))(\mathrm{Re}(p) - \mathrm{Im}(p))\\ &= \mathrm{Re}^2(p) - \langle \mathrm{Im}(p), \mathrm{Im}(p) \rangle\\ &= \mathrm{Re}^2(p) + |\mathrm{Im}(p)|^2\\ &= |p|^2,\end{aligned}$$

já que $\mathrm{Re}(\mathbb{H}) \perp \mathrm{Im}(\mathbb{H})$.

Por fim, assim como em $\mathbb{C}$ temos que $|pq| = |p| \cdot |q|$, quaisquer que sejam $p, q \in \mathbb{H}$. De fato, note que

$$|pq|^2 = (pq)(\overline{pq}) = p \underbrace{q\overline{q}}_{\in \mathbb{R}} \overline{p} = p\overline{p}q\overline{q} = |p|^2 \cdot |q|^2$$

e o resultado segue.

Agora, recorde que podemos pensar $\mathbb{C}$ como um espaço vetorial real de dimensão 2 com base $\{1, i\}$. Sendo assim, podemos fixar $z \in \mathbb{C}$ e definir

$$L_z : \mathbb{C} \to \mathbb{C}$$

por $L_z(w) = zw$ para cada $w \in \mathbb{C}$ (ou seja, uma multiplicação à esquerda). Daí, se $z = a + ib$ com $a, b \in \mathbb{R}$, temos

$$\begin{aligned} L_z(1) &= z \cdot 1 = a + ib = (a, b), \\ L_z(i) &= z \cdot i = -b + ia = (-b, a) \end{aligned}$$

de modo que podemos escrever

$$[L_z] = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

Agora, podemos definir o isomorfismo

$$\mathbb{C} \to gl(2, \mathbb{R})$$

que mapeia z em $[L_z]$. Segue daí que podemos pensar em números complexos como matrizes da forma $[L_z]$.

Assim como antes, vamos buscar estender esses argumentos para $\mathbb{H}$. Aqui, vamos pensar $\mathbb{H}$ como um espaço vetorial **complexo** de dimensão 2 com base $\{1, j\}$. De fato, podemos fazer isso pois dado $(z, w) \in \mathbb{C}^2$ com $z = (a, b)$ e $w = (\alpha, \beta)$, temos

$$z + wj = a + ib + (\alpha + i\beta)j = a + bi + \alpha j + \beta k$$

e, claramente, qualquer quatérnio q pode ser escrito na forma $z + wj$ (basta escolher as partes real e imaginária de q como acima).

Transposta esta barreira, defina, para $q = z + wj$ com $(z, w) \in \mathbb{C}^2$,

$$L_q : \mathbb{H} \to \mathbb{H}$$

por $L_q(v) = v \cdot q$ para cada $v \in \mathbb{H}$. Então L_q é uma transformação linear, pois, sendo $c \in \mathbb{C}$ e $u, v \in \mathbb{H}$ quaisquer, temos

$$(cu + v)q = (cu)q + (vq) = c(uq) + (vq).$$

Observe que a escolha de multiplicação à direita é feita pois estamos vendo $\mathbb{H}$ como um $\mathbb{C}$-espaço "à esquerda", em outras palavras, q não comuta com c em geral. Vale ressaltar que aqui $(\cdot)$ simboliza o produto *quaterniônico*. Agora, note que

$$\begin{aligned} L_q(1) &= 1 \cdot q = z + wj = (z, w), \\ L_q(j) &= j \cdot q = jz + jwj = -\overline{w} + \overline{z}j = (-\overline{w}, \overline{z}), \end{aligned}$$

de modo que

$$[L_q] = \begin{pmatrix} z & -\overline{w} \\ w & \overline{z} \end{pmatrix}.$$

E agora, podemos definir o isomorfismo

$$\mathbb{H} \to gl(2, \mathbb{C})$$

que mapeia q em $[L_q]$. Segue daí que podemos pensar em quatérnios como matrizes da forma $[L_q]$. Note a semelhança gritante com o caso complexo!

Observação. *Para o cálculo de $L_q(j)$, utilizamos a propriedade de que $zj = j\overline{z}$, $\forall z \in \mathbb{C}$, sendo j unidade imaginária quaterniônica. De fato, escrevendo $z = a + ib, a, b \in \mathbb{R}$ temos*

$$zj = (a + ib)j = aj + ibj = ja + bij = ja - bji = ja - jib = j(a - ib) = j\overline{z}.$$

Feita toda esta construção dos quatérnios, podemos agora falar de um de seus aspectos mais interessantes: a facilidade que ele nos dá para descrever rotações no espaço tridimensional.

Então, tome $q \in \mathbb{H}^*$ e defina, para todo $v \in \mathbb{H}$,

$$C_q : v \mapsto qvq^{-1},$$

chamada **conjugação por** q. Vamos ver que C_q nos dá uma rotação. Antes da verificação propriamente dita, temos alguns indícios que nos sugerem que, de fato, C_q é rotação. Primeiro, note que C_q não é a identidade e que q e 1 ficam fixos sob C_q.

Agora, note que dado $v \in \mathbb{H}$ qualquer, temos

$$|qvq^{-1}| = |q||v||q^{-1}| = |q||v||q|^{-1} = |v|,$$

de modo que podemos tomar, s.p.g., $q \in S^3 \subseteq \mathbb{H}$. Recorde, agora, que $\mathrm{Re}(\mathbb{H}) \perp \mathrm{Im}(\mathbb{H})$, de modo que o complemento ortogonal de 1 é $1^\perp = \mathrm{Im}(\mathbb{H})$. Consequentemente, como $C_q(1) = 1$, então $C_q(1^\perp) = 1^\perp = \mathrm{Im}(\mathbb{H})$. Esta última afirmação se deve ao seguinte teorema.

Teorema 4.3. *Se T é uma transformação linear ortogonal tal que $T(W) \subseteq W$, então $T(W^\perp) \subseteq W^\perp$.*

Demonstração. Note que, como T é ortogonal, então $T^{-1} = T^t = T^*$. Além disso, como $T(W) \subseteq W$ segue que $T(W) = W$ e, portanto, $W = T^{-1}(W)$, ou seja, T^{-1} deixa W invariante. Logo, dado $v \in W^\perp$, temos

$$0 = \langle v, T^*(w)\rangle = \langle v, T^{-1}(w)\rangle = \langle T(v), w\rangle,\ w \in W$$

ou seja, $T(W^\perp) = W^\perp$. □

Voltando, temos então que C_q deixa $\mathrm{Im}(\mathbb{H})$ invariante. Agora, como $q \in S^3 \subseteq \mathbb{H}$, segue que existem $\theta \in \mathbb{R}$ e $u \in \mathbb{H}$ tais que

$$q = \cos\theta + u\sin\theta = e^{u\theta},\ \text{com } u^2 = -1.$$

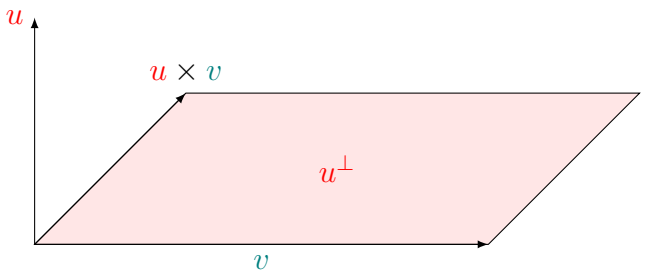
Lembrando que $C_q(q) = q$, segue que $C_q(u) = e^{u\theta}e^{u\pi/2}e^{-u\theta} = u \in \mathrm{Im}\mathbb{H}$. Então, dado $v \in u^\perp \subseteq \mathrm{Im}(\mathbb{H})$, temos

$$\begin{aligned}
C_q(v) &= e^{u\theta}ve^{-u\theta} \\
&= (\cos\theta + u\sin\theta)v(\cos\theta - u\sin\theta) \\
&= (\cos\theta v + uv\sin\theta)(\cos\theta + u\sin\theta) \\
&= \cos^2\theta v + uv\sin\theta\cos\theta - \sin\theta\cos\theta vu - \sin^2\theta uvu \\
&= (\cos^2\theta - \sin^2\theta)v + 2\sin\theta\cos\theta uv \\
&= \cos 2\theta v + u \times v\sin 2\theta \\
&= \text{ rotação de } 2\theta \text{ no plano } u^\perp \subseteq \mathrm{Im}(\mathbb{H}),
\end{aligned}$$

em que a quinta igualdade se deve ao fato de que

$$uv = u \times v - \langle u, v \rangle = u \times v, \text{ pois } v \in u^{\perp},$$
$$uvu = (uv)u = (u \times v)u = v,$$

como ilustrado abaixo.



Em suma, a construção que fizemos foi: C_q tem como conjuntos invariantes as partes real e imaginária de $\mathbb{H}$ e o subespaço gerado por u, que é imaginário puro. Então, podemos fazer a identificação $\mathbb{R}^3 \simeq \text{Im}(\mathbb{H})$ e considerar a restrição de C_q à parte imaginária (isso é possível porque ela é invariante). Ora, C_q, então, funciona como uma rotação, pois preserva norma, fixa um eixo e gira um plano ortogonal a este eixo.

A partir daí, podemos definir uma aplicação

$$\begin{array}{rccc} \varphi & : & S^3 \to \text{SO}(3) \\ & & q \mapsto C_q \end{array},$$

que leva q na rotação correspondente à conjugação por q. Note que de fato φ vai em $\text{SO}(3)$, já que C_q fixa um eixo e gira um plano, preservando a orientação. Também podemos ver que φ é um homomorfismo de grupos: $C_{qp} = C_q \circ C_p$. Pelo Teorema de Euler, φ é sobrejetiva. Além disso, é imediato que $\{\pm 1\} \subseteq \ker(\varphi)$. Note também que os únicos reais em $\ker(\varphi)$ são 1 e -1. Suponha, então, que exista $q \in \mathbb{H}$ não real tal que $q \in \ker(\varphi)$. Escrevendo $q = e^{u\theta}$, temos

$$qvq^{-1} = v, \ \forall v \in u^{\perp} \iff uv \sin\theta = vu \sin\theta \iff u \times v = v \times u \iff u \times v = 0.$$

Mas isso é absurdo, pois $v \in u^{\perp}$. Logo, $\ker(\varphi) = \{\pm 1\}$. Portanto, como $\ker(\varphi) \trianglelefteq S^3$ e φ é sobrejetora, segue que

$$S^3 \big/ \{\pm 1\} \simeq \text{SO}(3).$$

Observação. *Os que conhecem um pouco de geometria projetiva vão perceber imediatamente que*

$$S^3 \big/ \{\pm 1\} \simeq \text{SO}(3) \simeq \mathbb{R}P^3,$$

já que estamos identificando antípodas de S^3!

Mostramos agora uma forma alternativa de se obter o núcleo de φ. É claro que $\{\pm 1\} \subseteq \ker(\varphi)$. Seja q um quatérnio unitário e suponha que C_q é a transformação identidade, então, para $v \in u^{\perp}$, temos $C_q(v) = qvq^{-1} = v$, ou seja, $qv = vq$ para todo $v \in u^{\perp}$. Mas v é imaginário puro, então, necessariamente, q comuta com todos os imaginários unitários. Portanto, $q \in \mathbb{R}$, ou seja, $q = \pm 1$ e temos $\{\pm 1\} = \ker(\varphi)$.

5 Transformações lineares e a adjunta de uma transformação linear

Uma função linear real f é muitas vezes referida como uma função que tem como seu gráfico uma reta ou um plano ou, de maneira geral, um hiperplano, ou seja, uma função dada por

$$f(x) = a_1x_1 + \cdots + a_nx_n + a,$$

com $a, a_1, \ldots, a_n \in \mathbb{R}$. No contexto de espaços vetoriais, tomamos uma definição ligeiramente distinta.

Definição 5.1. *Sejam V e W espaços vetoriais sobre um corpo $\mathbb{F}$. Dizemos que uma transformação $T : V \to W$ é linear se, para todo $c \in \mathbb{F}$ e todos $\alpha, \beta \in V$, vale*

$$T(c\alpha + \beta) = cT\alpha + T\beta.$$

Note então que a função f descrita acima é linear quando $a = 0$.

Se $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ é uma base de V, então dado $\alpha \in V$, podemos escrevê-lo de forma única como

$$\alpha = k_1\alpha_1 + \cdots + k_n\alpha_n.$$

Dizemos que a $n-$upla $(k_1, \ldots, k_n)$ são as coordenadas de α relativas à base $\mathcal{B}$. É natural então o isomorfismo $V \to \mathbb{F}^n$ dado por

$$\alpha = k_1\alpha_1 + \cdots + k_n\alpha_n \mapsto (k_1, \ldots, k_n).$$

Observe ainda que uma matriz $m \times n$ tem a ação de levar um vetor de $\mathbb{F}^n$ num vetor de $\mathbb{F}^m$. É natural, então, pensar que toda transformação linear pode ser descrita por meio de uma matriz. De fato, não é difícil determinar esta matriz: basta notar que a ação de uma transformação linear T é totalmente determinada se soubermos como ela age sobre a base, e isto é um indicativo de que a representação matricial da transformação muda conforme a base muda.

Sejam V e W espaços vetoriais de dimensões n e m, respectivamente, com bases $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ e $\mathcal{B}' = \{\beta_1, \ldots, \beta_m\}$. Seja $T : V \to W$ uma transformação linear. Escreva, para cada $1 \leq j \leq n$,

$$T\alpha_j = \sum_{i=1}^{m} A_{ij}\beta_i. \qquad (*)$$

Então

$$\begin{pmatrix} A_{1j} \\ \vdots \\ A_{mj} \end{pmatrix}$$

são as coordenadas de $T\alpha_j$ na base $\mathcal{B}'$. Seja agora $\alpha \in V$, com coordenadas $(x_1, \ldots, x_n) =$

$[\alpha]_{\mathcal{B}}$. Então

$$\begin{aligned} T\alpha &= \sum_{j=1}^{n} x_j T\alpha_j \\ &= \sum_{j=1}^{n} x_j \sum_{i=1}^{m} A_{ij}\beta_i \\ &= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} x_j A_{ij} \right) \beta_i. \end{aligned}$$

Segue que $A = (A)_{ij}$ é a matriz que representa a transformação linear T e $[T\alpha]_{\mathcal{B}'} = A[\alpha]_{\mathcal{B}}$. A equação em $(*)$ diz apenas que $[T\alpha]_{\mathcal{B}'} = A_j$, a j-ésima coluna de A. Denotamos $A = [T]_{\mathcal{B}}^{\mathcal{B}'}$.

Consideremos então a composta de transformações lineares. Sejam V, W e Z espaços vetoriais de dimensões n, m e p, respectivamente. Sejam $T : V \to W$ e $L : W \to Z$ transformações lineares com matrizes A e B, respectivamente. Então a transformação composta $LT = L \circ T : V \to Z$ é também linear e dada por

$$\begin{aligned} (L \circ T)(c\alpha + \beta) &= L(T(c\alpha + \beta)) \\ &= L(cT\alpha + T\beta) \\ &= cLT\alpha + LT\beta. \end{aligned}$$

Vamos considerar como se expressa a matriz da composta LT. Sejam $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$, $\mathcal{B}' = \{\beta_1, \ldots, \beta_m\}$ e $\mathcal{B}'' = \{\gamma_1, \ldots, \gamma_p\}$ as bases de V, W e Z, respectivamente. Das discussões anteriores, temos

$$\begin{aligned} [T\alpha]_{\mathcal{B}'} &= A[\alpha]_{\mathcal{B}}, \\ L(T\alpha)]_{\mathcal{B}''} &= B[T\alpha]_{\mathcal{B}'}, \end{aligned}$$

logo,

$$[L(T\alpha)]_{\mathcal{B}''} = BA[\alpha]_{\mathcal{B}}.$$

Portanto, a matriz da transformação composta é simplesmente o produto $C = BA$. Daí, dado $\alpha \in V$, temos

$$T\alpha_j = \sum_{i=1}^{m} A_{ij}\beta_i \quad \text{e} \quad L\beta_j = \sum_{n=1}^{p} B_{nj}\gamma_n.$$

Então,

$$\begin{aligned} (LT)\alpha_j = L(T\alpha_j) &= L\left(\sum_{k=1}^{m} A_{kj}\beta_k \right) \\ &= \sum_{k=1}^{m} A_{kj}(L\beta_k) \\ &= \sum_{k=1}^{m} A_{kj} \left(\sum_{i=1}^{p} B_{ik}\gamma_i \right) \\ &= \sum_{i=1}^{p} \left(\sum_{k=1}^{m} B_{ik}A_{kj} \right) \gamma_i \end{aligned}$$

então, necessariamente, $C_{ij} = \sum_{k=1}^{m} B_{ik}A_{kj}$ e daqui segue a familiar definição de produto de matrizes.

Seguindo nesta discussão sobre coordenadas, surge um conceito que é importantíssimo e bastante útil: o de espaço dual. Definindo a soma de transformações lineares para as transformações $T, U : V \to W$ de modo natural por

$$(T+U)\alpha = T\alpha + U\alpha \quad \text{e} \quad (cT)\alpha = c(T\alpha),$$

então

$$\begin{aligned}(cT+U)(k\alpha+\beta) &= (cT)(k\alpha+\beta) + U(k\alpha+\beta)\\ &= cT(k\alpha+\beta) + U(k\alpha+\beta)\\ &= c[kT\alpha + T\beta] + (kU\alpha + U\beta)\\ &= k(cT)\alpha + kU\alpha + cT\beta + U\beta\\ &= k[(cT)+U]\alpha + [(cT)+U]\beta,\end{aligned}$$

o que nos diz que $cT + U$ é linear e, portanto, o conjunto de transformações lineares de V para W forma um espaço vetorial, que denotaremos por $L(V, W)$.

O espaço dual é, então, definido como $V^* = L(V, \mathbb{F})$. Fazem parte deste conjunto as funções da forma

$$f(c\alpha + \beta) = cf(\alpha) + f(\beta),$$

onde $\alpha, \beta \in V$ e $c, f(\alpha)$ e $f(\beta)$ são valores em $\mathbb{F}$. As funções f são chamados funcionais lineares.

Seja $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ uma base de V. Pode-se mostrar que existem únicas funções f_i, com $1 \leq i \leq n$, tais que

$$f_i(\alpha_j) = \delta_{ij},$$

onde $\delta_{ij} = 1$ se $i = j$ e é zero caso contrário (este é o *delta de Kronecker*). Desta forma, temos o seguinte resultado.

Teorema 5.1. *O conjunto $\{f_1, \ldots, f_n\}$ é uma base do espaço dual V^*. Portanto, ele tem dimensão finita n.*

Demonstração. De início, vamos mostrar que os f_i determinam um conjunto linearmente independente. Suponha que $f = k_1 f_1 + \cdots + k_n f_n = 0$, então $k_i = f(\alpha_i) = 0$ para todo i. Logo, o conjunto é L.I.. Se $f \in V^*$ e $V \ni \alpha = k_1\alpha_1 + \cdots + k_n\alpha_n$, veja que

$$f(\alpha) = \sum_{i=1}^{n} k_i f(\alpha_i) = \sum_{i=1}^{n} f(\alpha_i) f_i(\alpha),$$

donde segue que

$$f = \sum_{i=1}^{n} f(\alpha_i) f_i$$

e o conjunto dos f_i gera V^*. □

Como consequência imediata desse teorema, temos o seguinte corolário.

Corolário 5.1.1. *Existe um isomorfismo de V em V^* dado por $\alpha_i \mapsto f_i$.*

Note que, no corolário anterior, fixamos uma base para estabelecer o resultado que queríamos. Por isso, dizemos que o isomorfismo assim definido não é natural ou canônico. Vamos ver agora alguns fatos interessantes que surgem ao considerarmos o conjunto $V^{**} = (V^*)^*$, i.e., o conjunto das funções $L : V^* \to \mathbb{F}$ que associa, a cada função em V^*, um escalar.

Considere as funções da forma

$$L_\alpha(f) = f(\alpha).$$

Veja que são lineares, já que

$$L_\alpha(cf + g) = (cf + g)(\alpha) = cf(\alpha) + g(\alpha) = cL_\alpha(f) + L_\alpha(g).$$

Além disso, o mapeamento $L_\alpha \mapsto \alpha$ é um isomorfismo de espaços vetoriais, pois

$$L_{c\alpha+\beta}(f) = f(c\alpha + \beta) = cL_\alpha(f) + L_\beta(f),$$

ou seja,

$$L_{c\alpha+\beta} = cL_\alpha + L_\beta.$$

Corolário 5.1.2. *V, V^* e V^{**} têm a mesma dimensão.*

O interessante a se observar neste isomorfismo entre V e V^{**} é que não fixamos nenhuma base (não tivemos que fazer nenhum tipo de escolha!) e, por isso, dizemos que este é um isomorfismo natural (ou canônico).

Quando consideramos transformações lineares $T : V \to W$ entre espaços vetoriais, há uma outra transformação linear que é induzida de forma natural entre os espaços V^* e W^*. Defina

$$f(\alpha) = g(T\alpha),$$

onde $g \in W^*$. Então T define uma relação que, a cada funcional g em W, associa um funcional f em V^*; denotemos esta regra por

$$f = T^*(g) \ (= g \circ T).$$

A transformação $T^* : W^* \to V^*$ é chamada **transformação adjunta** de T.

Vamos encontrar uma relação bem interessante entre as matrizes das transformações T e T^*. Seja

$$A = (A_{ij})$$

a matriz da transformação T relativa às bases $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ e $\mathcal{B}' = \{\beta_1, \ldots, \beta_m\}$ de V e W respectivamente. Considere $\{f_i\}$ e $\{g_i\}$ as bases dos duais V^* e W^*, respectivamente. Como vimos acima,

$$A[\alpha_j]_{\mathcal{B}} = A_j,$$

a j-ésima coluna de A, donde segue que

$$f_i(T\alpha_j) = A_{ij}.$$

Denote por $B = (B_{ij})$ a matriz correspondente a T^*. Por definição,

$$T\alpha_j = \sum_{i=1}^{m} A_{ij}\beta_i \ \text{ e } \ T^*g_j = \sum_{i=1}^{n} B_{ij}f_i.$$

Além disso,

$$\begin{aligned}(T^*g_j)(\alpha_i) &= g_j(T\alpha_i)\\ &= g_j\left(\sum_{k=1}^{m} A_{ki}\beta_k\right)\\ &= \sum_{k=1}^{m} A_{ki}g_j(\beta_k)\\ &= \sum_{k=1}^{m} A_{ki}\delta_{jk}\\ &= A_{ji}.\end{aligned}$$

Segue que

$$T^*g_j = \sum_{i=1}^{n} A_{ji}f_i = \sum_{i=1}^{n} B_{ij}f_i$$

e, portanto, $A_{ji} = B_{ij}$. Assim, a matriz de T^* é, simplesmente, a transposta da matriz de T.

Para tornar tudo mais concreto, tomemos $V = \mathbb{R}^n$ com a base canônica $\{e_j\}$. Denote por $\{e_i^*\}$ a base de $(\mathbb{R}^n)^*$ dual a esta, i.e., aquela definida por $e_i^*(e_j) = \delta_{ij}$. Podemos identificar os vetores de $\mathbb{R}^n$, segundo essa base, por vetores coluna em $\mathbb{R}^{n\times 1}$ da forma

$$(x_1, \ldots, x_n) \longleftrightarrow \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Lembre que os elementos de $(\mathbb{R}^n)^*$ são transformações lineares de $\mathbb{R}^n$ em $\mathbb{R}$. Então, em termos de coordenadas, temos, para um funcional f qualquer em $(\mathbb{R}^n)^*$

$$f = \sum_{j=1}^{n} A_{1j}e_j^* = \sum_{i=1}^{n} f(e_i)e_i^*$$

e podemos identificar então

$$(f(e_1), \ldots, f(e_n)) \longleftrightarrow \begin{pmatrix} f(e_1) & \cdots & f(e_n) \end{pmatrix} \in \mathbb{R}^{1\times n}.$$

Procedendo como antes e lembrando que a representação matricial é feita com coordenadas, temos

$$\begin{aligned}[f][\alpha]_S &= \begin{pmatrix} f(e_1) & \cdots & f(e_n) \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\\ &= x_1f(e_1) + \cdots + x_nf(e_n)\\ &= f(x_1e_1 + \cdots x_ne_n)\\ &= f(\alpha),\end{aligned}$$

onde S denota a base canônica e $\alpha = x_1e_1 + \cdots + x_ne_n$. Nesta identificação, fixadas as bases do $\mathbb{R}^n$ e do seu dual, uma transformação linear do dual é simplesmente o produto

escalar. Isto pode parecer abstrato, mas, na verdade, é bem intuitivo para quem já trabalhou com vetores. Ainda utilizando a base canônica do $\mathbb{R}^n$, quando queremos saber a primeira coordenada de um dado vetor α, o que fazemos é exatamente calcular $\langle e_1, \alpha \rangle$. O isomorfismo "não natural" do qual falamos entre V e V^* aqui se aplica como $e_i^* \leftrightarrow e_i$ ou, se preferir,

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix} \longleftrightarrow \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

(no caso $i = 1$) com as identificações que fizemos anteriormente.

Consideremos agora esta correspondência $\mathbb{R}^k \simeq (\mathbb{R}^k)^*$. Seja $T : \mathbb{R}^n \to \mathbb{R}^m$ uma transformação linear e $T^* : (\mathbb{R}^m)^* \to (\mathbb{R}^n)^*$ sua adjunta. Ao mostrarmos que $[T^*] = [T]^t$, o que fizemos foi, em essência, mostrar que o seguinte diagrama comuta.

$$\begin{array}{ccc} (\mathbb{R}^m)^* & \xrightarrow{T^*} & (\mathbb{R}^n)^* \\ \wr\uparrow & & \wr\uparrow \\ \mathbb{R}^m & \xrightarrow{T^t} & \mathbb{R}^n \end{array}$$

6 Anuladores, Hiperespaços e Dimensão

Nesta seção, vamos estabelecer conexões entre funcionais lineares e subespaços de um espaço vetorial; em especial com um subespaço especial chamado hiperespaço. Além disso, vamos discutir sobre dimensão usando as ideias que desenvolvemos antes sobre identificações e quocientes.

Para um espaço vetorial de dimensão finita, a sua dimensão é o número de vetores que há em uma base. Quando consideramos transformações de um espaço vetorial em outro, podemos nos perguntar que tipo de relações há entre as respectivas dimensões desses espaços, se o conjunto imagem tem alguma propriedade interessante ou se o núcleo nos dá alguma informação de valor. Começamos com o seguinte teorema.

Teorema 6.1. *Sejam V e W espaços vetoriais sobre $\mathbb{F}$ e $T : V \to W$ uma transformação linear. A imagem de T, $T(V)$, é um subespaço de W e o núcleo* $\ker T$ *é um subespaço de V.*

Demonstração. Como T é sobrejetora sobre sua imagem, dados $\alpha', \beta' \in T(V) \subset W$, temos que existem $\alpha, \beta \in V$ tais que $\alpha' = T\alpha$ e $\beta' = T\beta$. Portanto, para todo $c \in \mathbb{F}$, temos

$$c\alpha' + \beta' = cT\alpha + T\beta = T(c\alpha + \beta).$$

Logo, $c\alpha' + \beta' \in T(V)$ e, portanto, $T(V)$ é um espaço vetorial.

Observando que, para todos $\alpha, \beta \in \ker T$ e $c \in \mathbb{F}$, temos da linearidade de T que $T(c\alpha + \beta) = 0$, segue o resultado. □

Vamos mostrar agora um dos teoremas mais úteis da álgebra linear.

Teorema 6.2 (Imagem e Núcleo)**.** *Sejam V e W espaços vetoriais de dimensão finita n e m, respectivamente, sobre $\mathbb{F}$ e $T : V \to W$ uma transformação linear. Então*

$$\dim T(V) + \dim \ker T = n.$$

Demonstração. Como vimos acima, temos que $\ker T$ é um subespaço de V. Então, seja $S = \{\alpha_1, \ldots, \alpha_k\}$ uma base de $\ker T$ e adicione um conjunto $S' = \{\alpha_{k+1}, \ldots, \alpha_n\}$ de $n-k$ vetores a esta base, de modo que $\{\alpha_1, \ldots, \alpha_n\}$ seja uma base de V. É claro que qualquer vetor $\alpha' \in T(V)$ deve ser imagem de uma combinação linear de vetores do conjunto S' (não de forma única, qualquer vetor de S poderia ser adicionado a esta combinação). Então $\{T\alpha_{k+1}, \ldots, T\alpha_n\}$ gera $T(V)$.

Vamos mostrar que $\{T\alpha_{k+1}, \ldots, T\alpha_n\}$ é um conjunto L.I.. Para tal, suponha que

$$x_{k+1}T\alpha_{k+1} + \cdots + x_nT\alpha_n = 0,$$

então

$$T(x_{k+1}\alpha_{k+1} + \cdots + x_n\alpha_n) = 0,$$

de modo que $x_{k+1}\alpha_{k+1} + \cdots + x_n\alpha_n \in \ker T$ e pode ser escrito como uma combinação linear dos vetores desta base, ou seja,

$$x_{k+1}\alpha_{k+1} + \cdots + x_n\alpha_n = x_1\alpha_1 + \cdots + x_k\alpha_k$$

e, da independência linear desses vetores, concluimos que $x_i = 0$ para todo i. Em particular, $x_{k+1} = \cdots = x_n$, logo, $\{T\alpha_{k+1}, ..., T\alpha_n\}$ é L.I..

Segue que

$$\dim T(V) + \dim \ker T = (n-k) + k = n = \dim V.$$

□

Vamos analisar agora este mesmo teorema no caso particular em que W é um subespaço de V. Vamos considerar inicialmente o que seria o quociente de V por W. Claro que isto não pode ser feito no sentido comum de divisão, o que faremos é usar as mesmas ideias que usamos na primeira seção deste texto, trabalhar com relações de equivalência. Defina

$$uRv \iff u - v \in W$$

para vetores $u, v \in V$.

Os casos mais interessantes de classes de equivalência aparecem quando consideramos os vetores que estão em V mas não estão em W. Isto porque W é espaço vetorial, logo, para todos $w, w' \in W$, temos $w - w' \in W$. No entanto, se $u, v \in V - W$, temos

$$u - v \in W \iff u = v + W.$$

Aqui, W denota um vetor qualquer de W. Pelo menos a ideia desta proposição é bem intuitiva e a prova não é difícil.

Definamos então

Definição 6.1. *O quociente* V/W *é o conjunto de vetores* $\{v + W : v \in V\}$.

Podemos definir operações neste conjunto de modo que ele seja um espaço vetorial sobre um corpo $\mathbb{F}$. Defina

Definição 6.2. *Dados* $u + W, v + W \in V/W$ *e* $c \in \mathbb{F}$*, defina*

$$(u + W) + (v + W) = (u + v) + W \ (= (u + v) + W + W),$$
$$c(u + W) = cu + W \ (= cu + cW).$$

Colocamos entre parênteses fatos intuitivos que demonstram a importância de W ser escolhido como um subespaço de V.

Antes de tudo, é necessário mostrar que esta operação está bem definida, isto é, tomando quaisquer representantes de duas classes de equivalência, o resultado obtido nas operações é sempre o mesmo. Mais especificamente, temos o seguinte teorema.

Teorema 6.3. *No conjunto quociente de V por W valem*

1. $v + W = v' + W \iff v - v' \in W$;
2. *a soma de dois elementos está bem definida;*
3. *a multiplicação por escalar está bem definida.*

Demonstração. 1. Suponha $v+W = v'+W$. Então $v \in v'+W$ implica que $v = v'+w$, com $w \in W$; portanto $v - v' = w \in W$.

Suponha que $v - v' \in W$. Como qualquer múltiplo de $v - v'$ também está em W (propriedade de espaço vetorial!), concluímos que $v \in v' + W$ e $v' \in v + W$; escreva $v = v' + w'$. Seja $x \in v + W$, então $x = v + \tilde{w} = v' + (w' + \tilde{w})$, para $\tilde{w} \in W$. Logo, $x \in v' + W$. Como estas são classes de equivalência com interseção não vazia, concluímos que $v + W = v' + W$.

2. Por definição, $(u+W)+(v+W) = (u+v)+W$. Seja então $u' \in u+W$ e $v' \in v+W$; portanto, $u' = u + w_1$ e $v' = v + w_2$. Segue que

$$(u' + W) + (v' + W) = (u' + v') + W = (u + v) + w_1 + w_2 + W = (u + v) + W.$$

3. Dado $c \in \mathbb{F}$, temos, por definição, $c(u + W) = cu + W$. Do mesmo modo, se $u' \in u + W$, então $u' = u + w_1$. Segue que

$$c(u' + W) = c(u + w_1) + W = cu + W.$$

Note que em todas as partes usamos que W é um espaço vetorial. □

Desta forma, pode-se verificar que V/W é, de fato, um espaço vetorial. Geometricamente, este espaço pode ser visto como o conjunto dos espaços obtidos ao transladar W por um certo vetor de V. Claro que, se o vetor de translação está em W, o efeito é nenhuma translação.

Considere $V = \mathbb{R}^3$ e W um plano do $\mathbb{R}^3$ pela origem (veremos a seguir por que esse tipo de conjunto é um subespaço). Ao considerarmos translações de W, as únicas translações que têm algum efeito são aquelas por vetores que não são paralelos a W (vetores que não pertencem a este subespaço). Todos os vetores que têm a mesma projeção na direção normal ao plano W, determinam a mesma translação sobre W, logo, entram na mesma classe de equivalência.

Veremos agora um teorema que permitirá a mesma conclusão que o Teorema 6.2, mas num caso particular.

Teorema 6.4. *Sejam V um espaço vetorial de dimensão n sobre um corpo $\mathbb{F}$ e W um subespaço de V. Então*

$$\dim V/W = \dim V - \dim W.$$

Demonstração. Seja $S_0 = \{w_1, \ldots, w_k\}$ uma base de W e selecione n vetores adicionais v_i tais que $S = \{w_1, \ldots, w_k, v_{k+1}, \ldots, v_{k+n}\}$ seja uma base de V. Vamos mostrar que $\{v_{k+1} + W, \ldots, v_n + W\}$ é um conjunto L.I. em V/W. Suponha que

$$(x_{k+1}v_{k+1} + W) + \cdots + (x_n v_n + W) = 0,$$

então

$$(x_{k+1}v_{k+1} + \cdots + x_{k+n}v_{k+n}) + W = 0.$$

Pelo modo como definimos o quociente, devemos ter que $x_{k+1}v_{k+1} + \cdots + x_{k+n}v_{k+n}$ é um vetor em W, logo,

$$x_{k+1}v_{k+1} + \cdots + x_{k+n}v_{k+n} = x_1 w_1 + \cdots + x_k w_k.$$

Como S é uma base de V, segue que $x_i = 0,\ i = 1, \ldots, k+n$. Em particular, são nulos para $k+1 \leq i \leq k+n$, donde segue que $\{v_{k+1} + W, \ldots, v_{k+n} + W\}$ é de fato L.I..

Como todo vetor de V é combinação dos vetores de $\{v_{k+1}, \ldots, v_{k+n}\}$ a menos de um vetor em W, concluímos que $\{v_{k+1} + W, \ldots, v_{k+n} + W\}$ gera V/W. Temos então

$$\dim {}^{V}\!/_{W} = n - k = \dim V - \dim W.$$

□

Quando se fala em transformações de modo geral, pode ser que elas não sejam bijetivas, ou mesmo injetivas ou sobrejetivas. Definindo uma transformação $T : V \to T(V)$, onde V e $T(V)$ são espaços vetoriais quaisquer, garantimos que T é sobrejetiva. Mas ainda há uma certa redundância na imagem de T, no sentido de que pode haver vetores que têm a mesma imagem por T. Neste caso, usando a linearidade de T, obtemos

$$T(u) = T(v) \iff T(u - v) = 0 \iff u - v \in \ker T.$$

Podemos proceder como fizemos antes e considerar as classes de equivalência $u + \ker T$ e o quociente $V/\ker T$. Definindo uma aplicação $L : {}^{V}\!/_{\ker T} \to T(V)$ por $L(u + \ker T) = T(u)$, vemos que ela herda a linearidade de T e que é bijetiva por sua construção. Portanto, ${}^{V}\!/_{\ker T}$ é isomorfo a $T(V)$ e temos $\dim {}^{V}\!/_{\ker T} = \dim T(V)$. Segue como consequência do teorema que provamos e deste isomorfismo o seguinte corolário.

Corolário 6.4.1. $\dim {}^{V}\!/_{\ker T} = \dim T(V) = \dim V - \dim \ker T.$

Recorde agora o seguinte fato conhecido: dados dois vetores não nulos α e β em $\mathbb{R}^n$, temos $\langle \alpha, \beta \rangle = 0$ só quando α é perpendicular a β. Desta forma, é natural definir subespaços de $\mathbb{R}^n$ utilizando o produto interno, i.e., definindo-os como o conjunto de vetores perpendiculares a uma certa direção.

Mas podemos fazer o mesmo para espaços sem produto interno? A resposta é sim e utilizaremos algo bem interessante que funciona como uma espécie de generalização do produto interno. Segue do Teorema 6.1 o seguinte resultado.

Lema 6.5. *Sejam V um espaço vetorial qualquer e V^* seu dual. Dado $f \in V^*$, o conjunto dos vetores em $\ker f$ é um subespaço de V. Além disso, se $f_1, f_2, \ldots, f_n \in V^*$, então $\bigcap_{i=1}^{n} \ker f_i$ é um subespaço de V.*

Se pudermos dar a noção de direção para um vetor em V, então a correspondência $\alpha_i \leftrightarrow f_i$ entre a base de v e de seu dual nos diz que um funcional f não se anula precisamente na direção de um vetor. Daí, poderíamos dizer, num certo sentido, que o conjunto $\ker f = \{\alpha \in V : f(\alpha) = 0\}$ é o conjunto dos vetores perpendiculares a um dado vetor de V. Daqui em diante, vamos detalhar mais esta ideia.

Definição 6.3. *Se V é um espaço vetorial sobre um corpo $\mathbb{F}$ e S é um subconjunto de V, então o anulador de S é o conjunto S^0 dos funcionais f em V^* tais que $f(\alpha) = 0$ para todos $\alpha \in V$.*

Sabendo que $(cf + g)(\alpha) = cf(\alpha) + g(\alpha)$, fica claro que S^0 é um subespaço de V^*. Usando a identificação $V \simeq V^{**}$ que estabelecemos antes, dada por $\alpha \leftrightarrow L_\alpha$, considere o anulador de S^0, i.e., o conjunto $S^{00} = \{L_\alpha : L_\alpha(f) = 0, f \in S^0\}$. Este conjunto é um espaço vetorial e, segundo a identificação, corresponde, em V, à intercessão dos núcleos de funcionais em S^0. Claro que $S \not\simeq S^{00}$, mas $S^{00} \simeq W$, onde W é o subespaço de V gerado por S. Resumimos no seguinte teorema:

Teorema 6.6. *Sejam V um espaço vetorial de dimensão finita sobre um corpo $\mathbb{F}$ e S um subconjunto de V. Então S^0 e S^{00} são subespaços de V^* e V^{**}, respectivamente. Além disso, com a identificação $V \simeq V^{**}$, S^{00} é o subespaço de V gerado por S.*

Demonstração. A primeira parte segue do parágrafo anterior. E que S^{00} é o subespaço de V gerado por S segue do próximo lema.

Lema 6.7. *Seja V um espaço vetorial de dimensão finita sobre $\mathbb{F}$ e W um subespaço de V. Então*

$$\dim W + \dim W^0 = \dim V.$$

Demonstração. Seja W de dimensão k e $\{\alpha_1, \ldots, \alpha_k\}$ uma base de W. Adicione $n - k$ vetores a este conjunto, de modo que $\{\alpha_1, \ldots, \alpha_n\}$ seja uma base de V. Construa então a base de V^* dual a esta $\{f_1, \ldots, f_n\}$ definida por

$$f_i(\alpha_j) = \delta_{ij}.$$

Então $\delta_{ij} = 0$ para $i \geq k + 1$ e $j \leq k$, de modo que $f_i \in W^0$ para $i \geq k + 1$.

Vamos mostrar que $\{f_{k+1}, \ldots, f_n\}$ é uma base de W^0. Este conjunto é L.I. por construção. Seja $f \in V^*$, então

$$f = \sum_{i=1}^{n} f(\alpha_i) f_i.$$

Se f está em W^0, então

$$f = \sum_{i=k+1}^{n} f(\alpha_i) f_i.$$

Logo, $\{f_{k+1}, \ldots, f_n\}$ é uma base de W^0, e está provado o lema. □

Usando este lema, temos as seguintes equações

$$\dim V = \dim W + \dim S^0,$$
$$\dim V = \dim S^0 + \dim S^{00}$$

e segue que $\dim S^{00} = \dim W$ e, portanto, $W \simeq S^{00}$. □

Vamos tratar agora de subespaços importantes e alguns resultados que nos darão informações sobre o isomorfismo entre V e V^*. Um subespaço N de um espaço vetorial V é chamado maximal se satisfaz a seguinte condição: se W é subespaço de V e $N \subseteq W$, então $W = N$ ou $W = V$. Assim, temos a definição a seguir.

Definição 6.4. *Um hiperespaço de um espaço vetorial V é um subespaço maximal próprio.*

Para começar, mostramos o seguinte fato.

Teorema 6.8. *Se f é um funcional não nulo em V, então* $\ker f$ *é um hiperespaço em V. Além disso, qualquer hiperespaço em V é o núcleo de algum funcional f (não necessariamente único!).*

Demonstração. Se f é não nulo, então existe $\alpha \in V$ tal que $f(\alpha) \neq 0$. Vamos mostrar que todo vetor de V é da forma $c\alpha + \gamma$, onde $\gamma \in \ker f$ e $c \in \mathbb{F}$. Seja $\beta \in V$ e defina

$$c = \frac{f(\beta)}{f(\alpha)},\ f(\alpha) \neq 0.$$

Então $f(c\alpha - \beta) = 0$, o que mostra que $\beta = c\alpha + \gamma$ para algum $\gamma \in \ker f$.

Além disso, essa representação é única, pois se $\beta = c'\alpha + \gamma'$ teríamos

$$(c - c')\alpha = \gamma' - \gamma.$$

Necessariamente, $c = c'$, pois $\alpha \notin \ker f$. O que mostramos é que, dado $\beta \in V$, existe um único escalar c tal que $\beta - c\alpha \in \ker f$. Então $N = \ker f$ é um hiperespaço. Denote $c = g(\beta)$, então g é um funcional em V tal que seu espaço nulo é N, um hiperespaço em V. □

O teorema anterior pode ser visto da seguinte forma: os hiperespaços em $\mathbb{R}^3$ são os planos pela origem. Seja α o vetor normal a um plano π e $f \in V^*$ que é não nulo apenas na direção de α (vimos que isto é possível de acordo com o isomorfismo que há entre V e V^* quando fixamos uma base). Qualquer vetor de $\mathbb{R}^3$ tem uma componente sobre o plano π e uma componente que é um múltiplo de α; digamos que $\beta = c\alpha + \gamma$, onde γ é um vetor paralelo ao plano π. Então, quando fazemos $\beta - c\alpha$, o que obtemos é um vetor paralelo a π, pois estamos retirando a componente de β na direção de α. Tente relacionar isto com o produto escalar!

Lema 6.9. *Sejam $f, g \in V^*$. Então g é um múltiplo escalar de f se, e somente se,* $\ker f \subseteq \ker g$.

Demonstração. Suponha $g = kf$. Então, se $\alpha \in \ker f$, temos $g(\alpha) = kf(\alpha) = 0$. Logo, $\ker f \subseteq \ker g$.

Suponha que $\ker f \subseteq \ker g$. Seja $\alpha \in V$ tal que $f(\alpha) \neq 0$ e denote

$$c = \frac{g(\alpha)}{f(\alpha)}.$$

Seja $h = g - cf$ um funcional. Observe que $\ker f$ é um hiperespaço em V e que h é nula em $\ker f$, já que $\ker f \subseteq \ker g$ e também é nula sobre os múltiplos de α pela definição de c. Portanto, $h \equiv 0$ e segue que $g = cf$. □

Sob a ótica do que consideramos para $\mathbb{R}^3$, vemos que, para funcionais f e g este lema significa que $g = cf$ se f e g se anulam sobre os vetores de um mesmo plano. Ou seja, neste caso, seus núcleos são os mesmos.

Teorema 6.10. *Sejam, $g, f_1, \ldots, f_r$ funcionais em V com núcleos $N, N_1, \ldots, N_r$ respectivamente. Então g é uma combinação linear de $f_1, \ldots, f_r$ se, e somente se, $\bigcap_{i=1}^{r} N_i \subseteq N$.*

Demonstração. Claramente, se $g = \sum_{i=1}^{r} c_i f_i$ e $\alpha \in V$ é tal que $f_i(\alpha) = 0$ para cada i, então $\bigcap_{i=1}^{r} N_i \subseteq N$.

Suponha então que $\bigcap_{i=1}^{r} N_i \subseteq N$. O teorema já foi mostrado para o caso $r = 1$ no lema anterior. Suponha que vale para $r = k - 1$. Sejam $g', f_1', \ldots, f_{k-1}'$ as restrições de $g, f_1, \ldots, f_{k-1}$ ao subespaço $N_k = \ker f_k$. Então $g', f_1', \ldots, f_{k-1}'$ são funcionais em N_k. Se $\alpha \in N_k$ é tal que $f_i(\alpha) = 0$, $i = 1, 2, \ldots, k-1$, então $\alpha \in \bigcap_{i=1}^{k} N_i$ e $g'(\alpha) = 0$ (fazemos esta hipótese, pois o caso $\bigcap_{i=1}^{r} N_i = \emptyset$ é degenerado, já que cada N_i é um hiperespaço).

Pela hipótese de indução, existem escalares c_i tais que

$$g' = \sum_{i=1}^{k-1} c_i f_i'. \qquad (\star)$$

Considere o funcional

$$h = g - \sum_{i=1}^{k-1} c_i f_i.$$

Então, por $(\star)$, temos que h é nulo sobre qualquer vetor de N_k (já que os funcionais em V coincidem com as restrições sobre $N_k \subset V$) e segue do lema que $h = c_k f_k$. Logo,

$$g = \sum_{i=1}^{k} c_i f_i.$$

□

7 Lema de Bézout e decomposição de Jordan

Vamos aqui introduzir o conceito de operadores e apresentar alguns teoremas e propriedades importantes a eles relacionados, como o lema de Bézout, e suas consequências. Mostraremos também algumas aplicações e relações com resolução de equações diferenciais parciais e também para a decomposição de Jordan de uma matriz. Aqui assumimos familiaridade com a noção de polinômios sobre um corpo $\mathbb{F}$.

Ademais, mostraremos várias propriedades interessantes da teoria dos polinômios sobre um corpo, e esperamos poder evidenciar a incrível e inegável semelhança (como o

próprio nome sugere) entre polinômios primos e números primos. De fato, muitas das propriedades de divisão satisfeitas pelos números são também válidas, com as devidas adaptações, para polinômios em $\mathbb{F}[x]$; e.g., um análogo do Teorema Fundamental da Aritmética é válido para polinômios.

Começamos então com uma definição.

Definição 7.1. *Uma transformação linear $T : V \to V$ é chamada operador linear.*

Dados um operador T de um espaço vetorial V sobre $\mathbb{F}$ e um polinômio $p(x) = a_0 + a_1x + \cdots + a_nx^n$ com coeficientes em $\mathbb{F}$, podemos definir $p(T) : V \to V$ por

$$p(T) = a_0 + a_1T + \cdots + a_nT^n.$$

Daí, o núcleo $\ker p(T) = \{v \in V \mid p(T)v = 0\}$ é *T-invariante*, i.e., invariante sob T. Isto significa que, se $v \in \ker p(T)$, então $Tv \in \ker p(T)$, pois T e $p(T)$ são operadores de V que comutam. Por exemplo, em $V = C^\infty(\mathbb{R})$ definimos o operador derivada $D : V \to V$, de modo que $\ker p(D) = \{y \ : \ p(D)y = 0\}$ nada mais é que o espaço solução de uma EDO de coeficientes constantes com polinômio característico p.

Outro exemplo, agora em dimensão finita, é o de um operador T de $V = \mathbb{R}^n$. Nesse caso, $\dim V = n$ e pode-se mostrar que existe um polinômio não nulo $p(x) \in \mathbb{R}[x]$ tal que $p(T) = 0$. De fato, definindo $\mathcal{V} = \hom(V, V)$ (o espaço dos homomorfismos de V em V), temos $\mathcal{V} = gl(n, \mathbb{R})$ e $\dim \mathcal{V} = n^2$. Daí, segue que

$$\mathcal{V} \ni a_0 + a_1T + a_2T^2 + \cdots + a_{n^2}T^{n^2} = 0$$

implica que nem todos os coeficientes a_i são nulos, já que a combinação acima é l.d.; este é o polinômio $p(x)$ procurado.

Daqui em diante passaremos a discutir polinômios em corpos, com o objetivo de apresentar o background necessário para tratar da decomposição de espaços. Não obstante, os assuntos sobre polinômios discutidos nas próximas páginas é belo e interessante em si mesmo, merecendo ser discutido aqui. A título de esclarecimento, $\mathbb{F}$ sempre denotará um corpo a menos que mencionado o contrário.

Lema 7.1. *Sejam f e d polinômios não nulos sobre $\mathbb{F}$ tais que $\deg d \leq \deg f$. Então existe um polinômio $g \in \mathbb{F}[x]$ tal que*

$$f - dg = 0 \quad \textit{ou} \quad \deg(f - dg) < \deg f.$$

Demonstração. Suponha que

$$f = a_mx^m + \sum_{i=0}^{m-1} a_ix^i, \ a_m \neq 0,$$
$$d = b_nx^n + \sum_{i=0}^{n-1} b_ix^i, \ b_n \neq 0.$$

Então $m \geq n$ e

$$f - \left(\frac{a_m}{b_n}\right) x^{m-n}d = 0 \qquad \text{ou} \qquad \deg\left[f - \left(\frac{a_m}{b_n}\right) x^{m-n}d\right] < \deg f,$$

de modo que podemos tomar $g = \left(\frac{a_m}{b_n}\right) x^{m-n}$. □

A partir desse lema pode-se mostrar que a familiar "divisão longa" de polinômios com coeficientes reais ou complexos é possível sobre qualquer corpo.

Teorema 7.2. *Se f, d são polinômios sobre $\mathbb{F}$ e d é não nulo, então existem $q, r \in \mathbb{F}[x]$ tais que*

(i) $f = dq + r$;

(ii) *ou* $r \equiv 0$ *ou* $\deg r < \deg d$.

Além disso, q e r satisfazendo (i) e (ii) são únicos.

Demonstração. Se $f \equiv 0$ ou $\deg f < \deg d$, basta tomar $q \equiv 0$ e $r = f$. Do contrário, i.e., $f \not\equiv 0$ e $\deg f \geq \deg d$, então o lema acima nos permite escolher um polinômio g tal que

$$f - dg = 0 \quad \text{ou} \quad \deg(f - dg) < \deg f.$$

Se $f - dg \not\equiv 0$ e $\deg(f - dg) \geq \deg d$, escolhemos h tal que

$$(f - dg) - dh = 0 \quad \text{ou} \quad \deg[f - d(g + h)] < \deg(f - dg).$$

Continuando desse modo, obteremos eventualmente polinômios q, r tais que

$$r \equiv 0 \quad \text{ou} \quad \deg r < \deg d \quad \text{e} \quad f = dq + r.$$

Agora, suponha que também tenhamos $f = dq_1 + r_1$, com $r_1 \equiv 0$ ou $\deg r_1 < \deg d$. Assim,

$$dq + r = dq_1 + r_1 \Longrightarrow d(q - q_1) = r_1 - r.$$

Se $q - q_1 \not\equiv 0$, então $d(q - q_1) \not\equiv 0$ e

$$\deg d + \deg(q - q_1) = \deg(r - r_1),$$

o que é absurdo pois $\deg(r - r_1) < \deg d$. Segue que $q - q_1 \equiv 0$ e $r - r_1 \equiv 0$. □

Vale ressaltar aqui a semelhança com o algoritmo da divisão de Euclides para inteiros. Aqui, a noção de polinômio "maior" ou "menor" é capturada pelo grau, deg. O próximo passo natural é definir a noção de divisibilidade para polinômios.

Definição 7.2. *Seja d um polinômio não nulo sobre $\mathbb{F}$. Se $f \in \mathbb{F}[x]$, o teorema anterior nos mostra que existe no máximo um polinômio $q \in \mathbb{F}[x]$ tal que $f = dq$. Se tal q existe, dizemos que d divide f, que f é divisível por d ou ainda que f é um múltiplo de d, e chamamos q o quociente de f e d. Também escrevemos $q = f/d$.*

Daí seguem alguns corolários conhecidos no Ensino Médio para polinômios de coeficientes reais.

Corolário 7.2.1. *Sejam $f \in \mathbb{F}[x]$ e $c \in \mathbb{F}$. Então f é divisível por $x - c \iff f(c) = 0$.*

Demonstração. Pelo teorema acima, $f = (x - c)q + r$, sendo r um polinômio constante. Ademais, temos

$$f(c) = 0q(c) + r(c) = r(c),$$

de modo que $r = 0 \iff f(c) = 0$. □

Corolário 7.2.2. *Um polinômio f de grau n sobre $\mathbb{F}$ tem no máximo n raízes em $\mathbb{F}$.*

Demonstração. Se $n = 0, 1$ o resultado é imediato. Assuma o resultado válido para polinômios de grau $n-1$. Se a é raiz de f, então $f = (x-a)q$ com $\deg q = n-1$. Como $f(b) = 0 \iff a = b$ ou $q(b) = 0$, segue por hipótese de indução que f tem no máximo n raízes. □

Esse corolário suscita uma discussão sobre raízes múltiplas. Para tal, as derivadas formais de um polinômio são úteis: a *derivada* de um polinômio

$$f = c_0 + c_1 x + \cdots + c_n x^n$$

é o polinômio

$$f' = c_1 + 2c_2 x + \cdots + n c_n x^{n-1}.$$

Também podemos usar a notação $Df = f'$, sendo D o operador derivada definido anteriormente (agora agindo em $\mathbb{F}[x]$). Também temos derivadas formais de ordem superior, como $f'' = D^2 f$, $f^{(3)} = D^3 f$ e assim por diante. Com essas ferramentas, podemos provar a fórmula de Taylor.

Teorema 7.3 (Fórmula de Taylor)**.** *Sejam $\mathbb{F}$ um corpo de característica 0*[†]*, $c \in \mathbb{F}$ e $n \in \mathbb{N}$. Se $f \in \mathbb{F}[x]$ com $\deg f \leq n$, então*

$$f = \sum_{k=0}^{n} \frac{(D^k f)}{k!}(c)(x-c)^k.$$

Demonstração. Pelo binômio de Newton, vem

$$x^m = [c + (x-c)]^m = \sum_{k=0}^{m} \binom{m}{k} c^{m-k}(x-c)^k = c^m + mc^{m-1}(x-c) + \cdots + (x-c)^m,$$

que é a fórmula de Taylor para $f = x^m$. Se

$$f = \sum_{m=0}^{n} a_m x^m,$$

então

$$(D^k f)(c) = \sum_m a_m (D^k x^m)(c)$$

e

$$\begin{aligned}
\sum_{k=0}^{n} \frac{(D^k f)(c)}{k!}(x-c)^k &= \sum_k \sum_m a_m \frac{(D^k x^m)(c)}{k!}(x-c)^k \\
&= \sum_m a_m \sum_k \frac{(D^k x^m)(c)}{k!}(x-c)^k \\
&= \sum_m a_m x^m \\
&= f.
\end{aligned}$$

□

[†]A característica de um corpo é o menor número natural n tal que adicionando a identidade multiplicativa n vezes obtemos a identidade aditiva. A característica é definida como zero se tal natural n não existe (exemplo: $\mathbb{R}$).

É interessante notar que como $1, (x-c), \ldots, (x-c)^n$ são linearmente independentes, a fórmula de Taylor nos dá o *único* método de escrever f como combinação linear de polinômios $(x-c)^k, 0 \leq k \leq n$. A título de curiosidade, a fórmula de Taylor ainda vale (com os devidos ajustes ao enunciado) para corpos de característica finita, nos quais podemos ter $k! = 0$.

Junto com raízes vem a discussão acerca de sua *multiplicidade*, que é o maior natural r tal que $(x-c)^r$ divide f. É evidente que a multiplicidade de uma raiz é menor ou igual que $\deg f$. Para polinômios sobre corpos de característica 0, a multiplicidade de uma raiz c de f está relacionada com o número de derivadas de f que se anulam em c, como mostra o teorema a seguir.

Teorema 7.4. *Seja $\mathbb{F}$ um corpo de característica 0 e f um polinômio sobre $\mathbb{F}$ com $\deg f \leq n$. Então $c \in \mathbb{F}$ é raiz de f com multiplicidade r se, e só se,*

$$(D^k f)(c) = 0,\ 0 \leq k \leq r-1 \qquad e \qquad (D^r f)(c) \neq 0.$$

Demonstração. Seja r a multiplicidade de c raiz de f. Então existe um polinômio g tal que $f = (x-c)^r g$ e $g(c) \neq 0$. De fato, do contrário teríamos f divisível por $(x-c)^{r+1}$ pelo primeiro dos dois corolários acima. Aplicando a fórmula de Taylor para g, segue que

$$f = (x-c)^r \left[\sum_{m=0}^{n-r} \frac{(D^m g)}{m!}(c)(x-c)^m \right] = \sum_{m=0}^{n-r} \frac{(D^m g)}{m!}(c)(x-c)^{r+m}.$$

Como vimos, f é escrita de maneira única como combinação linear de $(x-c)^k,\ 0 \leq k \leq n$. Daí, segue que

$$\frac{(D^k f)}{k!}(c) = \begin{cases} 0,\ 0 \leq k \leq r-1 \\ \dfrac{(D^{k-r} g)}{(k-r)!}(c),\ r \leq k \leq n. \end{cases}$$

Portanto, $(D^k f)(c) = 0$ para $0 \leq k \leq r-1$ e $(D^r f)(c) = g(c) \neq 0$. Reciprocamente, se essas condições são satisfeitas então segue da fórmula de Taylor que existe um polinômio g tal que $f = (x-c)^r g$ e $g(c) \neq 0$. Por fim, suponha que r não seja o maior inteiro positivo tal que $(x-c)^r$ divide f. Então existe um polinômio h tal que $f = (x-c)^{r+1} h$, o que implica em $g = (x-c)h$ pelo segundo corolário acima. Daí, $g(c) = 0$, absurdo. □

Agora, para dar prosseguimento à nossa discussão sobre polinômios e continuar a tecer as semelhanças dessa teoria com a Teoria dos Números, introduzimos o conceito de ideal, necessário para a definição de mdc de polinômios.

Definição 7.3. *Um ideal em $\mathbb{F}[x]$ é um subespaço M de $\mathbb{F}[x]$ tal que dados $f \in \mathbb{F}[x], g \in M$ temos $fg \in M$.*

Dessa definição segue que $M = d\mathbb{F}[x]$ é ideal em $\mathbb{F}[x]$, sendo d um polinômio sobre $\mathbb{F}$. De fato, M é não vazio; em particular, contém d. Dados $f, g \in \mathbb{F}[x]$ e $c \in \mathbb{F}$, temos

$$c(df) - dg = d(cf - g) \in M,$$

de modo que M é subespaço. Por fim, M contém $(df)g = d(fg)$. O ideal M é chamado *ideal principal gerado por d.*

Ademais, dados $d_1, \ldots, d_n$ polinômios sobre $\mathbb{F}$, a soma M dos ideais $d_i\mathbb{F}[x]$ é também um ideal. De fato, dado $p \in M$ existem $f_1, \ldots, f_n \in \mathbb{F}[x]$ tais que

$$p = d_1 f_1 + \cdots + d_n f_n.$$

Se g é um polinômio qualquer sobre $\mathbb{F}[x]$, então

$$gp = pg = d_1(f_1 g) + \cdots + d_n(f_n g),$$

de modo que $gp \in M$ e M é ideal, chamado *ideal gerado* por $d_1, \ldots, d_n$.

Teorema 7.5. *Dado M ideal não nulo em $\mathbb{F}[x]$, existe um único polinômio mônico $d \in \mathbb{F}[x]$ tal que M é o ideal principal gerado por d. Em outras, palavras, todo ideal de $\mathbb{F}[x]$ é principal.*

Demonstração. Por hipótese, M contém um polinômio não nulo. Tomemos, entre todos os polinômios não nulos de M, aquele que seja mônico e de menor grau, d. Podemos escolher d mônico, pois do contrário bastaria escolher um polinômio de grau mínimo e multiplicá-lo por um escalar para torná-lo mônico. Agora, dado $f \in M$ podemos escrever

$$f = dq + r,$$

com $r \equiv 0$ ou $\deg r < \deg d$. Como $d \in M$, temos $dq \in M$ e $f - dq = r \in M$. Como d tem grau minimal, não podemos ter $\deg r < \deg d$, de modo que $r \equiv 0$ e $M = d\mathbb{F}[x]$.

Suponha g um polinômio mônico tal que $M = g\mathbb{F}[x]$. Daí, existem polinômios f, g não nulos tais que $d = gp$ e $g = dq$, donde segue que $d = dpq$ e

$$\deg d = \deg d + \deg p + \deg q.$$

Segue que $\deg p = 0 = \deg q$ e, como d, g são mônicos, $p = 1 = q$. Portanto, $d = g$. □

Vale notar que na demonstração acima usamos um caso especial de um fato mais geral e útil: dados p polinômio não nulo em um ideal M e $f \in M$ não divisível por p, então $f = pq + r$ onde o "resto" r pertence a M, é não nulo e tem grau menor que p. Em princípio, é sempre possível encontrar o polinômio mônico gerador de um ideal não nulo, pois podemos obter um polinômio, no ideal, de grau minimal após uma quantidade finita de divisões sucessivas.

Corolário 7.5.1. *Dados $p_1, \ldots, p_n$ polinômios não todos nulos sobre $\mathbb{F}$, existe um único polinômio mônico $d \in \mathbb{F}[x]$ tal que*

(a) d está no ideal gerado por $p_1, \ldots, p_n$;

(b) d divide cada um dos p_i.

Todo polinômio que satisfaça (a) e (b) também satisfaz

(c) d é divisível por todo polinômio que divida cada um dos p_i.

Demonstração. Seja d o gerador mônico do ideal

$$M = p_1\mathbb{F}[x] + \cdots + p_n\mathbb{F}[x].$$

Todo elemento de M é divisível por d, logo cada um dos p_i também o é. Suponha que f seja um polinômio que divide cada um dos p_i. Então existem polinômios $g_1, \ldots, g_n$ tais que $p_i = fg_i,\ 1 \leq i \leq n$. Além disso, como $d \in M$ então existem polinômios $q_1, \ldots, q_n$ tais que

$$d = p_1 q_1 + \cdots + p_n q_n,$$

donde

$$d = f[g_1q_1 + \cdots + g_nq_n].$$

Mostramos que d satisfaz (a), (b) e (c). Agora, se d' é qualquer polinômio satisfazendo (a) e (b), segue de (a) e da definição de d que d' é um múltiplo escalar de d e satisfaz (c) também. Por fim, se d' é mônico então $d' = d$. □

É inegável a semelhança de d definido na demonstração acima com o mdc. De fato, tratemos de defini-lo agora.

Definição 7.4. *Dados $p_1, \ldots, p_n$ polinômios não todos nulos sobre $\mathbb{F}$, o gerador mônico d do ideal*

$$p_1\mathbb{F}[x] + \cdots + p_n\mathbb{F}[x]$$

*é o maior divisor comum (*mdc*) de $p_1, \ldots, p_n$. Dizemos que $p_1, \ldots, p_n$ são relativamente primos ou primos entre si se o seu* mdc *é 1 ou, equivalentemente, se o ideal por eles gerado é $\mathbb{F}[x]$.*

Por exemplo, se considerarmos o corpo $\mathbb{C}$, então $\text{mdc}(x + 2, x^2 + 8x + 16) = 1$. De fato, o ideal M gerado por esses dois polinômios contém

$$x^2 + 8x + 16 - x(x + 2) = 6x + 16,$$

de modo que ele também contém

$$6x + 16 - 6(x + 2) = 4.$$

Logo, o polinômio 1 pertence a M e temos $M = \mathbb{C}$.

Agora, vamos demonstrar o teorema análogo, para polinômios, do Teorema Fundamental da Aritmética, i.e., que todo polinômio pode ser escrito (de maneira única a menos de ordem) como produto de polinômios "primos". Essa fatoração nos dá uma ferramenta eficiente para encontrar o mdc de uma quantidade finita de polinômios e, em particular, nos dá uma maneira efetiva de verificar se dados polinômios são relativamente primos ou não.

Definição 7.5. *Um polinômio $f \in \mathbb{F}[x]$ é dito redutível sobre $\mathbb{F}$ se existem $g, h \in \mathbb{F}[x]$ de grau ≥ 1 tais que $f = gh$ e, do contrário, f é dito irredutível sobre $\mathbb{F}$. Um polinômio irredutível não constante sobre $\mathbb{F}$ é chamado polinômio primo sobre $\mathbb{F}$, e podemos também dizer que ele é um primo em $\mathbb{F}[x]$.*

Note então que $x^2 + 1$ é redutível sobre $\mathbb{C}$ mas é primo sobre $\mathbb{R}$. Começamos agora introduzindo o primeiro teorema necessário para provar o nosso teorema sobre a fatoração de polinômios.

Teorema 7.6. *Sejam $p, f, g \in \mathbb{F}[x]$. Suponha que p é primo e que p divide fg. Então p divide f ou p divide g.*

Demonstração. Sem perda de generalidade, assuma p mônico. Como p é primo, os únicos divisores mônicos de p são 1 e p. Seja $d = \text{mdc}(f, p)$. Então, temos $d = 1$ ou $d = p$, já que d é um polinômio mônico que divide p. Se $d = p$, então p divide f e terminamos. Do contrário, existem f_0, p_0 tais que

$$1 = f_0f + p_0p \implies g = f_0fg + p_0pg = (fg)f_0 + p(p_0g).$$

Como p divide fg, ele divide $(fg)f_0$, e claro que p divide $p(p_0g)$. Logo, p divide g. □

Corolário 7.6.1. *Se p é primo e divide $\prod_{i=1}^{n} f_i$, então p divide um dos f_i.*

Demonstração. Procedemos por indução. O caso $n = 2$ foi tratado no teorema anterior. Suponha que o corolário vale para $n = k$ e que p divide $\prod_{i=1}^{k+1} f_i$. Como p divide $f_{k+1}\left(\prod_{i=1}^{k} f_i\right)$, então p divide f_{k+1} ou p divide $\prod_{i=1}^{k} f_i$. Por hipótese de indução, se p divide $\prod_{i=1}^{k} f_i$ então p divide f_j para algum $1 \leq j \leq k$, e segue que p divide f_i para algum $1 \leq i \leq k+1$. □

Podemos, agora, demonstrar nosso tão aguardado resultado.

Teorema 7.7. *Todo polinômio não constante em $\mathbb{F}[x]$ pode ser fatorado como produto de primos mônicos em $\mathbb{F}[x]$ em uma e, a menos de ordem, apenas uma maneira.*

Demonstração. Suponha f não constante sobre $\mathbb{F}$. Como polinômios de grau 1 são irredutíveis, terminamos se $\deg f = 1$. Suponha, então, $\deg f \geq 2$. Por indução, assuma que o teorema vale para todos os polinômios mônicos não constantes de grau menor que n. Se f é irredutível, então ele já está fatorado como produto de primos mônicos; do contrário, $f = gh$ com g e h mônicos não constantes de grau $< n$. Logo, por hipótese de indução, g e h podem ser fatorados como produtos de primos mônicos em $\mathbb{F}[x]$ e f, por conseguinte, também pode.

Agora, a unicidade: suponha que

$$f = \prod_{i=1}^{m} p_i = \prod_{j=1}^{n} q_j,$$

com $p_1, \ldots, p_m$ e $q_1, \ldots, q_n$ primos mônicos sobre $\mathbb{F}$. Então p_m divide o produto dos q_j, i.e., divide algum dos q_j. Como os q_j e os p_i são primos mônicos, temos

$$q_i = p_m. \tag{$*$}$$

Daí, segue que se $m = 1$ ou $n = 1$ teremos $m = 1 = n$, pois

$$\deg f = \sum_{i=1}^{m} \deg p_i + \sum_{j=1}^{n} \deg q_j,$$

e nesse caso terminamos. Suponhamos então $m > 1$ e $n > 1$. Reordenando os q_j, podemos assumir $p_m = q_n$, de modo que

$$p_1 \cdots p_{m-1} p_m = q_1 \cdots q_{n-1} p_m,$$

donde segue que

$$p_1 \cdots p_{m-1} = q_1 \cdots q_{n-1}.$$

Por hipótese de indução, segue que $q_1 \cdots q_{n-1}$ é apenas uma reordenação de $p_1 \cdots p_{m-1}$. Esse fato junto com ($*$) nos mostra que a fatoração de f como produto de primos mônicos é única a menos de ordem. □

Note que, analogamente aos inteiros, na fatoração acima do polinômio mônico não constante f podemos repetir algum dos fatores primos. Se $p_1, \ldots, p_r$ são os primos mônicos *distintos* que aparecem na fatoração de f, então

$$f = p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r},$$

sendo os expoentes n_i a quantidade de vezes que p_i aparece na fatoração de f. Note que essa decomposição também é claramente única (a menos de ordem); ela é chamada *decomposição prima* de f. Pode-se verificar que todo divisor mônico de f tem a forma

$$p_1^{m_1} p_2^{m_2} \cdots p_r^{m_r},\ 0 \leq m_i \leq n_i. \tag{$**$}$$

De ($**$), segue que o mdc de uma quantidade finita de polinômios mônicos não constantes $f_1, \ldots, f_n$ é obtido combinando todos os primos mônicos que ocorrem simultaneamente nas fatorações de $f_1, \ldots, f_n$. O expoente ao qual cada primo é elevado é o maior para o qual a respectiva potência prima é fator de cada f_i; se não houver nenhuma potência prima não trivial, os polinômios são relativamente primos. Em símbolos, se

$$f = p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$$

e

$$g = p_1^{m_1} p_2^{m_2} \cdots p_r^{m_r},$$

então

$$\operatorname{mdc}(f, g) = \prod_{i=1}^{r} p_i^{\alpha_i}, \quad \text{com } \alpha_i = \min\{n_r, m_r\}.$$

Para finalizar a discussão sobre polinômios, apresentamos alguns teoremas e uma discussão sobre corpos algebricamente fechados e a relação disso com raízes de polinômios em corpos.

Teorema 7.8. *Seja f um polinômio mônico não constante sobre $\mathbb{F}$ e seja*

$$f = p_1^{n_1} \cdots p_k^{n_k}$$

a fatoração prima de f. Para cada $1 \leq j \leq k$, faça

$$f_j = f/p_j^{n_j} = \prod_{i \neq j} p_i^{n_i}.$$

Então $f_1, \ldots, f_k$ são primos entre si.

Demonstração. Basta notar que como

$$\begin{aligned} f_1 &= p_2^{n_2} p_3^{n_3} \cdots p_r^{n_r}, \\ f_2 &= p_1^{n_1} p_3^{n_3} \cdots p_r^{n_r}, \\ &\vdots \\ f_k &= p_1^{n_1} p_2^{n_2} \cdots p_{r-1}^{n_{r-1}}, \end{aligned}$$

então segue que

$$\operatorname{mdc}(f_1, \ldots, f_n) = \prod_{i=1}^{k} p_i^{\min\{0, n_i\}} = 1.$$

□

Teorema 7.9. *Seja f um polinômio sobre $\mathbb{F}$ com derivada f'. Então f é produto de polinômios irredutíveis distintos sobre $\mathbb{F}$ se, e só se, f e f' são primos entre si.*

Demonstração. Suponha que na fatoração prima de f sobre $\mathbb{F}$ tenhamos um polinômio primo não constante p repetido. Então $f = p^2 h$ para algum $h \in \mathbb{F}[x]$ e, portanto,

$$f' = p^2 h' + 2pp'h$$

e p divide f'. Logo, $\text{mdc}(f, f') \neq 1$.

Agora, suponha que

$$f = p_1 \cdots p_k,$$

com $p_1, \ldots, p_k$ polinômios irredutíveis não constantes distintos sobre $\mathbb{F}$. Faça $f_j = f/p_j$. Então, derivando, temos

$$f' = p_1' f_1 + \cdots + p_k' f_k.$$

Seja p um polinômio primo que divide f e f'. Então $p = p_i$ para algum i. Ora, mas p_i divide f_j para $i \neq j$, e como p_i também divide

$$f' = \sum_{j=1}^{k} p_j' f_j,$$

segue que p_i divide $p_i' f_i$. Logo, p_i divide p_i' ou f_i. Mas p_i não divide f_i, pois $p_1, \ldots, p_k$ são distintos; portanto, p_i divide p_i', o que é absurdo, pois $\deg p_i' < \deg p_i$. Logo, nenhum primo divide f e f', i.e., $\text{mdc}(f, f') = 1$. □

Definição 7.6. *O corpo $\mathbb{F}$ é dito algebricamente fechado se todo polinômio primo sobre $\mathbb{F}$ tem grau 1.*

Então, dizer que $\mathbb{F}$ é algebricamente fechado significa que todo polinômio mônico irredutível sobre $\mathbb{F}$ é da forma $x - c$, i.e., um monômio. Já observamos que todo monômio é irredutível, seja qual for $\mathbb{F}$. Uma definição equivalente de corpo algebricamente fechado é um corpo $\mathbb{F}$ tal que todo polinômio não constante $f \in \mathbb{F}[x]$ pode ser escrito na forma

$$f = c(x - c_1)^{n_1} \cdots (x - c_k)^{n_k},$$

com c escalar, $c_1, \ldots, c_k \in \mathbb{F}$ distintos e $n_1, \ldots, n_k \in \mathbb{N}$. Uma terceira forma de definir seria dizer que se f é um polinômio não constante sobre $\mathbb{F}$, então existe $c \in \mathbb{F}$ tal que $f(c) = 0$.

Essa forma equivalente de definir corpos algebricamente fechados evidencia a conexão dessa definição com a existência de raízes de f em $\mathbb{F}$. Segue daí que o corpo $\mathbb{R}$ dos números reais **não é** algebricamente fechado, já que $x^2 + 1$ é irredutível de grau 2 sobre $\mathbb{R}$ ou, equivalentemente, porque $x^2 + 1$ não tem raiz real. O famoso Teorema Fundamental da Álgebra afirma que o corpo $\mathbb{C}$ dos números complexos é algebricamente fechado; ele também torna evidente as possibilidades de fatoração para polinômios com coeficientes reais. De fato, se f é um polinômio com coeficientes reais e c é uma raiz complexa de f, então o conjugado complexo $\overline{c}$ de c também é raiz de f. Portanto, as raízes imaginárias de f devem ocorrer em pares (de conjugados), de modo que o conjunto das raízes de f tem a forma

$$\left\{t_1, \ldots, t_k, c_1, \overline{c_1}, \ldots, c_r, \overline{c_r}\right\},$$

com $t_1, \ldots, t_k \in \mathbb{R}$ e $c_1, \ldots, c_r \in \mathbb{C}$ têm parte imaginária não nula. Segue que f é fatorado como

$$f = c(x - t_1) \cdots (x - t_k) p_1 \cdots p_r,$$

onde p_i é o polinômio quadrático

$$p_i = (x - c_i)(x - \overline{c_i}).$$

Observe que p_i tem coeficientes reais. De fato, escrevendo $c = \alpha + \beta i,\ \alpha, \beta \in \mathbb{R}$, temos

$$p_i = x^2 - x(\alpha + \beta i + \alpha - \beta i) + \alpha^2 + \beta^2 = x^2 - 2\alpha x + \alpha^2 + \beta^2.$$

Concluímos que todo polinômio irredutível sobre $\mathbb{R}$ tem grau 1 ou 2. Cada polinômio sobre $\mathbb{R}$ é produto de certos fatores lineares, obtidos a partir das raízes reais de f, e certos polinômios quadráticos irredutíveis, obtidos a partir das raízes imaginárias de f.

8 Teorema da Decomposição Primária

Nesta seção vamos tratar do assunto de decomposições de espaços e como isso pode ser útil com uma boa escolha de espaços. As ideias apresentadas sobre transformações lineares e polinômios são todas integradas aqui num teorema bastante bonito que é o Teorema da Decomposição Primária. Ao final da seção apresentamos uma aplicação desse teorema à solução de equações diferenciais lineares.

Seja V um espaço vetorial de dimensão finita e $B = \{\alpha_1, \ldots, \alpha_n\}$ uma base de V. Podemos escrever V como

$$V = W_1 + \cdots + W_n,$$

onde cada W_i é o espaço gerado por $\{\alpha_i\}$. Essa soma de espaços tem uma característica especial que é $W_i \cap W_j = \{0\}$ se $i \neq j$ e, se β_i é um vetor em W_i, $i = 1, \ldots, n$, então

$$\beta_1 + \cdots + \beta_n = 0 \Longrightarrow \beta_1 = \cdots = \beta_n = 0,$$

já que B é uma base de V. Este é um caso especial de algo mais geral que definimos a seguir.

Definição 8.1. *Seja V um espaço vetorial de dimensão finita e $W_1, \ldots, W_k$ subespaços de V. Dizemos que eles são independentes se*

$$\beta_1 + \cdots + \beta_k = 0 \Longrightarrow \beta_1 = \cdots = \beta_k = 0,$$

com $\beta_i \in W_i$, $i = 1, \ldots, k$. Se V se escreve como soma desses subespaços, denotamos

$$V = W_1 \oplus \cdots \oplus W_k = \bigoplus_{i=1}^{k} W_i,$$

e dizemos que V é a soma direta desses subespaços.

Seja $T : V \to V$ um operador linear em V e W um subespaço de V. A ideia de invariância também se aplica a subespaços de V. Dizemos que W é *invariante por T* se $T(W) \subseteq W$. Se conseguirmos achar uma decomposição de V em soma direta de subespaços vetoriais invariantes por T, seremos capazes de simplificar enormemente a matriz em que T se expressa.

Digamos que $V = W_1 \oplus \cdots \oplus W_k$, onde cada W_i é invariante por T. Sejam B_i, $i = 1, \ldots, k$, bases dos respectivos W_i. Então $B = B_1 \cup \cdots \cup B_k$ é uma base de V. Nesta base, T tem a forma simplificada

$$[T]_B = \begin{pmatrix} \star & \star & & & & & & \\ \star & \star & & & & & & \\ & & \star & \star & & & & \\ & & \star & \star & & & & \\ & & & & \ddots & & & \\ & & & & & \star & \star & \star \\ & & & & & \star & \star & \star \\ & & & & & \star & \star & \star \end{pmatrix},$$

onde os espaços vazios contêm zeros e as estrelas indicam entradas não necessariamente nulas. Em particular, se cada W_i fosse unidimensional, teríamos

$$[T]_B = \begin{pmatrix} \lambda & & & & & & & \\ & \lambda & & & & & & \\ & & \lambda & & & & & \\ & & & \lambda & & & & \\ & & & & \ddots & & & \\ & & & & & \lambda & & \\ & & & & & & \lambda & \\ & & & & & & & \lambda \end{pmatrix},$$

onde os λ não são necessariamente iguais, ou seja, T seria diagonalizável. Um exemplo simples onde isso sempre acontece: podemos decompor V como soma direta da imagem e do núcleo de um operador T (que são claramente invariantes por T).

Seja $f = a_n x^n + \cdots + a_1 x + a_0$. Definimos o valor de f aplicado ao operador T por

$$f(T) = a_n T^n + \cdots + a_1 T + a_0 I.$$

Observe que $f(T)$ é também um operador linear em T, basta mostrar que T^k é linear para todo $k \in \mathbb{N}$. Dizemos que f anula o operador T se $f(T) = 0$, ou seja, se $f(T)$ é o operador nulo em V. Um polinômio com tal propriedade sempre existe, como foi visto anteriormente. Um tal polinômio é da forma

$$a_{n^2} T^{n^2} + \cdots + a_1 T + a_0 I,$$

já que a dimensão do espaço de operadores lineares em V tem dimensão n^2 se V tem dimensão n.

O interessante é que o conjunto dos anuladores de um operador T é um ideal no espaço de polinômios.

Teorema 8.1. *Seja $M = \{f \in \mathbb{F}[x] : f(T) = 0\}$. Então M é um ideal.*

Demonstração. Se $g, f \in M$, então

$$(cg + f)(T) = cg(T) + f(T) = 0,$$

logo, M é subespaço de $\mathbb{F}[x]$. Se h é qualquer polinômio, então

$$(hf)(T) = h(T)f(T) = 0,$$

e $hf \in M$. □

Definição 8.2. *O gerador mônico deste ideal é chamado polinômio minimal de T.*

Já temos o suficiente para enunciar o teorema. Seja T um operador linear em V de dimensão finita. A ideia principal do teorema é decompor um espaço V de dimensão finita como soma de núcleos de polinômios em T. Mais precisamente, seja

$$p = p_1^{r_1} \cdots p_k^{r_k}$$

o polinômio minimal de T, decomposto em fatores irredutíveis. Sabemos que $p(T)$ é o operador nulo em V, então temos $V = \ker p(T)$. Isto não é de forma alguma revelador. Buscaremos mostrar que V pode ser decomposto como soma direta dos subespaços $\ker p_i^{r_i}(T)$.

Teorema 8.2 (Teorema da Decomposição Primária – TDP)**.** *Seja T um operador linear num espaço de dimensão finita V sobre o corpo $\mathbb{F}$. Seja p o polinômio minimal de T,*

$$p = p_1^{r_1} \cdots p_k^{r_k},$$

onde p_i são polinômios distintos irredutíveis sobre $\mathbb{F}$. Seja $W_i = \ker p_i^{r_i}(T)$. Então

- *$V = W_1 \oplus \cdots \oplus W_k$;*
- *Cada W_i é invariante por T;*
- *Se $T_i = T|_{W_i}$, o polinômio minimal de T_i é $p_i^{r_i}$.*

Demonstração. Sejam os polinômios

$$f_i = \frac{p}{p_i^{r_i}} = \prod_{j \neq i} p_j^{r_j}.$$

Vimos anteriormente que eles são coprimos, então existem polinômios $g_1, \ldots, g_k$ tais que

$$1 = g_1 f_1 + \cdots + g_k f_k.$$

Denote $h_i = g_i f_i$ e $E_i = h_i(T)$. Segue que

$$I = E_1 + \cdots + E_k.$$

Note que o polinômio minimal divide $h_i h_j$ sempre que $i \neq j$, então $h_i(T)h_j(T) = 0$. Temos diretamente que

$$E_i E_j = 0 \qquad \text{e} \qquad E_i^2 = E_i,$$

basta usar a equação acima obtida para I.

Vamos mostrar agora que $W_i = E_i(V)$. Seja $\alpha \in W_i$, então $E_j \alpha = 0$ se $j \neq i$ por construção dos E_i. Segue que

$$\alpha = (E_1 + \cdots + E_k)\alpha = E_i \alpha,$$

logo, $\alpha \in E_i(V)$ e vem $W_i \subseteq E_i(V)$. Seja agora $\alpha \in E_i(V)$, então $\alpha = E_i \beta$ para algum $\beta \in V$. Segue que

$$\begin{aligned} p_i^{r_i}(T)\alpha &= p_i^{r_i}(T) E_i \beta \\ &= p_i^{r_i} h_i(T) \beta \\ &= 0, \end{aligned}$$

pois p divide $p_i h_i$. Segue que $\alpha \in W_i$ e temos $W_i = E_i(V)$.

Observe que escrevemos I como a soma das projeções nos espaços W_i e, além disso, todo vetor de V deve estar no núcleo de algum $p_i^{r_i}(T)$. O fato dos W_i serem independentes segue do fato de que $E_i E_j = 0$ com $i \neq j$. Segue que

$$V = W_1 \oplus \cdots \oplus W_k.$$

Agora, T comuta com qualquer polinômio em T, então, se $\alpha \in W_i$, temos

$$p_i^{r_i}(T)T\alpha = Tp_i^{r_i}(T)\alpha = T0 = 0,$$

logo $T\alpha \in W_i$. Portanto, cada W_i é invariante por T.

Se $T_i = T|_{W_i}$ é a restrição de T a W_i, por construção,

$$p_i^{r_i}(T_i) = 0.$$

Seja g qualquer tal que $g(T_i) = 0$, então

$$g(T)f_i(T) = 0.$$

Logo, $p = p_i^{r_i} f_i$ divide $g f_i$, ou seja, $p_i^{r_i}$ divide g. Então $p_i^{r_i}$ é o polinômio minimal de T_i. □

Consideramos agora equações diferenciais lineares da forma

$$\frac{d^n f}{dx^n} + a_{n-1}\frac{d^{n-1} f}{dx^{n-1}} + \cdots + a_1 \frac{df}{dx} + a_0 f = 0, \qquad (*)$$

onde cada a_i é constante. Observe que, se $p = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$, podemos reescrever esta equação diferencial na forma

$$p(D)f = 0,$$

onde D é o operador de derivação no espaço das funções infinitamente deriváveis C^∞.

Queremos aplicar o teorema que acabamos de provar ao espaço

$$V = \{f \in C^\infty : p(D)f = 0\},$$

ou seja, o espaço solução da equação diferencial $(*)$. Observe que não temos, a priori, que V tem dimensão finita, mas utilizaremos aqui o fato de que o TDP vale para dimensão qualquer. Veja que o polinômio minimal de D nesse espaço é p; suponha que ele fatore como

$$p = p_1^{r_1} \cdots p_k^{r_k}.$$

Aplicando o TDP, o nosso problema se reduz a resolver as equações

$$p_i^{r_i}(D)f = 0, \quad i = 1, \ldots, k.$$

Suponha agora que trabalhamos no corpo dos números complexos $\mathbb{C}$. Então todo polinômio se fatora como produto de fatores lineares da forma $x - \alpha$, com $\alpha \in \mathbb{C}$. Nosso problema se reduz a resolver

$$(D - \alpha)^r f = 0.$$

Teorema 8.3. *Para todo r natural tem-se*

$$(D-\alpha)^r f = e^{\alpha t} D^r (e^{-\alpha t} f).$$

Demonstração. Para $r = 1$, temos

$$\begin{aligned}(D-\alpha)f &= e^{\alpha t}e^{-\alpha t}(Df - \alpha f)\\ &= e^{\alpha t}(e^{-\alpha t}Df - \alpha e^{-\alpha t}f)\\ &= e^{\alpha t}D(e^{-\alpha t}f).\end{aligned}$$

Suponha que vale para r qualquer, então

$$(D-\alpha)^r f = e^{\alpha t} D^r (e^{-\alpha t} f).$$

Segue

$$\begin{aligned}(D-\alpha)^{r+1} f &= (D-\alpha)[e^{\alpha t}D^r(e^{-\alpha t}f)]\\ &= \alpha e^{\alpha t}D^r(e^{-\alpha t}f) + e^{\alpha t}D^{r+1}(e^{-\alpha t}f) - \alpha e^{\alpha t}D^r(e^{-\alpha t}f)\\ &= e^{\alpha t}D^{r+1}(e^{-\alpha t}f).\end{aligned}$$

□

Então segue que

$$(D-\alpha)^r f = 0 \iff e^{\alpha t}D^r(e^{-\alpha t}f) = 0,$$

e obtemos, por integração,

$$e^{-\alpha t}f = a_{r-1}t^{r-1} + \cdots + a_1 t + a_0,$$

ou seja,

$$\begin{aligned}f(t) &= e^{\alpha t}(a_{r-1}t^{r-1} + \cdots + a_1 t + a_0)\\ &= q(t)e^{\alpha t},\end{aligned}$$

com $\text{grau}(q) = r - 1$.

Observe que os componentes

$$e^{\alpha t}, te^{\alpha t}, \ldots, t^{r-1}e^{\alpha t}$$

são vetores linearmente independentes (ja que $e^{\alpha t} \neq 0$), logo, formam um espaço de dimensão r.

Voltemos agora para a solução de $p(D)f = 0$. Suponha que

$$p = (x-\alpha_1)^{r_1}\cdots(x-\alpha_k)^{r_k},$$

onde os α_i são as raízes distintas de p. Então a solução geral da EDO é da forma

$$f(t) = q_1(t)e^{\alpha_1 t} + \cdots + q_k(t)e^{\alpha_k t},$$

onde $p_i(t)$ é um polinômio em t de grau $r_i - 1$.

O TDP garante diretamente que

$$q_1(t)e^{\alpha_1 t}, \ldots, q_k(t)e^{\alpha_k t}$$

são independentes. A dimensão de V é

$$\begin{aligned}\sum_{i=1}^{k}(\text{grau}(p_i)+1) &= \sum_{i=1}^{k}((r_i-1)+1)\\ &= \text{grau}(p)\\ &= n.\end{aligned}$$

Segue então que a solução geral da equação diferencial se expressa como combinação linear de n soluções específicas de $p(D)f = 0$.

Para a discussão que segue, definimos

Definição 8.3. *Um operador N num espaço vetorial V é nilpotente se existe um inteiro positivo r tal que $N^r = 0$.*

Seja T um operador num espaço vetorial V e suponha que seu polinômio minimal p fatore como produto de fatores lineares, i.e.,

$$p = (x-c_1)^{r_1}\cdots(x-c_k)^{r_k},$$

onde $c_1,\ldots,c_k$ são os autovalores distintos de T. Como na demonstração do TDP, temos

$$I = E_1 + \cdots + E_k$$

e

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

onde $W_i = \ker(T-c_i)^{r_i}$ e E_i é a projeção em W_i. Seja $D = c_1E_1 + \cdots + c_kE_k$, então D é diagonalizável.

Teorema 8.4. *Seja T um operador num espaço vetorial V de dimensão finita sobre um corpo $\mathbb{F}$. Suponha que o polinômio minimal de T se escreve como produto de fatores lineares. Então T se escreve de forma única como $T = D + N$, onde D é diagonalizável e N é nilpotente. Além disso, $DN = ND$.*

Demonstração. Seja D como definido anteriormente e denote

$$N = T - D = (T-c_1I)E_1 + \cdots + (T-c_kI)E_k.$$

Como cada projeção é um polinômio em T, então T comuta com cada projeção, do que segue que

$$N^r = (T-c_1I)^rE_1 + \cdots + (T-c_kI)^rE_k = 0,$$

onde $r = \max\{r_1,\ldots,r_k\}$, logo, N é nilpotente e temos $T = D + N$. Agora, ambos N e D são polinômios em T, então comutam.

Para provar a unicidade, suponha que $T = D' + N'$ com D' diagonalizável e N' nilpotente. Então

$$D - D' = N' - N.$$

Note que, para um R suficientemente grande,

$$(N'-N)^R = \sum_{i=0}^{R}\binom{R}{i}(-1)^i(N')^{R-i}N^i = 0.$$

Donde vem que $D-D'$ deve ser nilpotente, o que só acontece se $D-D' = 0$, já que ambos são diagonais. Logo, $D = D'$ e $N = N'$. □

9 O teorema de Cayley-Hamilton

A demonstração mais comum e provavelmente mais conhecida do Teorema de Cayley-Hamilton é algébrica e usando "força bruta". Nesta seção, mostraremos o bonito Teorema Fundamental das Funções Simétricas, importante no estudo da Teoria de Galois [1], e daremos uma demonstração topológica do Teorema de Cayley-Hamilton, favorita dos analistas.

É curioso que a primeira versão desse teorema foi provada por William Hamilton para os quatérnios, em 1853. A versão "atual" apareceu pela primeira vez em um *memoir* de Arthur Cayley que introduzia o conceito de álgebra matricial, em 1858; nele, Cayley provou o teorema para o caso $n = 2$ e afirmou que o havia verificado para $n = 3$, mas que

"I have not thought it necessary to undertake the labour of a formal proof of the theorem in the general case of a matrix of any degree."

Em tradução livre,

"Não julguei necessário me dar ao trabalho de fazer uma prova formal do teorema no caso geral de uma matriz de qualquer grau."

O caso geral do teorema foi mostrado pela primeira vez por Ferdinand Frobenius em 1878, 25 anos depois de Hamilton enunciar a primeira versão! Passemos então às definições necessárias para enunciar o Teorema Fundamental das Funções Simétricas – TFFS.

Definição 9.1. *Um polinômio $P = P(x_1, \ldots, x_n)$ em n variáveis é dito simétrico se, dada qualquer permutação τ dos índices, $P(x_{\tau(1)}, \ldots, x_{\tau(n)}) = P$.*

Exemplo 9.1. *Os polinômios*

$$p_k = x_1^k + \cdots + x_n^k$$

*são simétricos para $1 \leq k \leq n$. Eles também são **homogêneos**, i.e., todos os termos não nulos têm mesmo grau.*

Definição 9.2 (Polinômios Simétricos Elementares – PSE)**.** *Os n polinômios simétricos elementares, $\sigma_1, \ldots, \sigma_n$, são definidos como:*

$$\begin{aligned}
\sigma_1 &= \sum_i x_i \\
\sigma_2 &= \sum_{i<j} x_i x_j \\
\sigma_3 &= \sum_{i<j<k} x_i x_j x_k \\
&\vdots \\
\sigma_n &= x_1 x_2 \cdots x_n.
\end{aligned}$$

Antes de prosseguir ao TFFS, é interessante observar o seguinte teorema, de Vieta, que nos dá uma primeira sugestão do TFFS.

Teorema 9.1 (Vieta)**.** *Seja $p(z)$ um polinômio mônico de grau n com raízes $r_1, \ldots, r_n$. Sejam $\sigma_1, \ldots, \sigma_n$ os n PSE nos r_i. Então*

$$p(z) = z^n - \sigma_1 z^{n-1} + \sigma_2 z^{n-2} + \cdots + (-1)^n \sigma_n.$$

Demonstração. Se expandirmos

$$\prod_{i=1}^{n}(x - r_i),$$

os termos serão precisamente $(-1)^{n-k} r_1^{b_1} \cdots r_n^{b_n} x^k$, com b_i igual a 1 ou 0 conforme r_i esteja ou não no produto, respectivamente, e k a quantidade de r_i excluídos, de modo que o número total de fatores no produto é n (contando x^k com multiplicidade k). Como há n escolhas binárias (incluir r_i ou x), temos 2^n parcelas. Agrupando os termos pelo grau obtemos os n polinômios simétricos elementares nos r_i. □

Exemplo 9.2. *Seja $T \in gl(n, \mathbb{C})$ com autovalores $\lambda_1, \ldots, \lambda_n$. Por Vieta,*

$$\det(\lambda I - T) = p_T(\lambda) = \lambda^n + \sum_{k=1}^{n}(-1)^k \sigma_k \lambda^{n-k},$$

sendo σ_k os PSE nos autovalores de T.

Teorema 9.2 (TFFS)**.** *Qualquer polinômio simétrico em n variáveis $x_1, \ldots, x_n$ pode ser escrito de forma única como um polinômio nos PSE $\sigma_1, \ldots, \sigma_n$.*

Demonstração. Apresentamos aqui a demonstração clássica, que remonta a Gauss.

Seja f o polinômio simétrico a ser representado. O conjunto dos termos de f com um dado grau é também um polinômio simétrico, e se pudermos representar cada um deles como um polinômio nos σ_i, podemos representar f; logo, podemos assumir s.p.g. que f é *homogêneo*.

Agora, ordene os termos de f *lexicograficamente*, i.e., defina

$$a x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n} > b x_1^{j_1} x_2^{j_2} \cdots x_n^{j_n}$$

se $i_1 > j_1$, ou se $i_1 = j_1$ e $i_2 > j_2$, ou se $i_1 = j_1, i_2 = j_2$ e $i_3 > j_3$, etc., e então ordene os termos de f de maneira decrescente.

Como f é simétrico, para cada termo $c x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}$ que f contiver, então f também conterá todos os termos possíveis que parecem com esse, mas com os expoentes permutados (seus "conjugados"). Segue então que o termo líder de f, digamos $c_1 x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}$, tem $i_1 \geq i_2 \geq \cdots \geq i_n$. Seja

$$g_1 = c_1 \sigma_1^{i_1 - i_2} \sigma_2^{i_2 - i_3} \cdots \sigma_{n-1}^{i_{n-1} - i_n} \sigma_n^{i_n}.$$

Temos que g_1 é simétrico com mesmo termo líder que f. Logo, $f - g_1$ é simétrico com um termo líder menor (lexicograficamente), $c_2 x_1^{j_1} x_2^{j_2} \cdots x_n^{j_n}$. Como antes, segue da simetria que $j_1 \geq j_2 \geq \cdots \geq j_n$. Logo, podemos fazer

$$g_2 = c_1 \sigma_1^{j_1 - j_2} \sigma_2^{j_2 - j_3} \cdots \sigma_{n-1}^{j_{n-1} - j_n} \sigma_n^{j_n},$$

de modo que g_2 tem o mesmo termo líder que $f - g_1$ e $f - g_1 - g_2$ tem termo líder que é menor ainda.

Continuemos de tal maneira. O algoritmo eventualmente termina sem termos restantes, pois há uma quantidade finita de monômios $x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}$ de um dado grau possíveis para o primeiro lugar. Portanto, em algum momento teremos

$$f - g_1 - g_2 - \cdots - g_k = 0 \iff f = g_1 + g_2 + \cdots + g_k,$$

que é a representação desejada de f como polinômio nos σ_i.

Para provar a unicidade, basta mostrar que o polinômio nulo em $x_1, \ldots, x_n$ é unicamente representado como o polinômio nulo em $\sigma_1, \ldots, \sigma_n$. Isso é verdade pois dois produtos distintos de polinômios elementares $\sigma^{k_1} \cdots \sigma^{k_n}$ não têm mesmo termo líder, já que o termo líder de $\sigma_1^{k_1} \cdots \sigma_n^{k_n}$ é

$$x_1^{k_1+\cdots+k_n} x_2^{k_2+\cdots+k_n} \cdots x_n^{k_n},$$

e a correspondência

$$(k_1, \ldots, k_n) \mapsto (k_1 + \cdots + k_n, \ldots, k_{n-1} + k_n, k_n)$$

é injetiva.

Logo, os termos líderes em uma soma de produtos disjuntos de polinômios simétricos elementares não se cancelam, e a soma só é nula se for vazia. □

A demonstração acima, apesar de clássica, não é a melhor, no sentido de que não se aproveita totalmente da simetria de f. Daremos uma demonstração alternativa, após enunciar e demonstrar o Teorema 9.5.

Teorema 9.3. *O conjunto das matrizes diagonalizáveis em $\mathbb{C}^n$ é denso.*

Demonstração. A ideia do argumento é encontrar um subconjunto do conjunto $\mathcal{D}$ das matrizes diagonalizáveis que seja denso. Consideremos então o conjunto

$$\mathcal{U} = \{T \in \mathrm{gl}(n, \mathbb{C}) : \lambda_i \neq \lambda_j \text{ se } i \neq j\} = \{T \in \mathrm{gl}(n, \mathbb{C}) : \Delta_T \neq 0\} \subset \mathcal{D},$$

o conjunto das matrizes com autovalores distintos, i.e., o conjunto das matrizes com

$$\Delta_T = \prod_{i \neq j} (\lambda_i - \lambda_j) \neq 0.$$

Ora, mas Δ_T nada mais é que uma função nas n variáveis $\lambda_1, \ldots, \lambda_n$:

$$\Delta_T = f(\lambda_1, \ldots, \lambda_n).$$

Note que em Δ_T temos tanto o fator $\lambda_i - \lambda_j$ quanto $\lambda_j - \lambda_i$. Assim, f é *simétrico*, e segue do Teorema 9.2 que podemos escrever

$$f(\lambda_1, \ldots, \lambda_n) = g(\sigma_1, \ldots, \sigma_n).$$

Do Teorema 9.2, segue que g é um polinômio nos coeficientes de p_T. Ora, mas p_T é um polinômio nas entradas de T, de modo que g é um polinômio nas entradas de T. Agora, supondo que g se anulasse em uma vizinhança de T teríamos, pela fórmula de Taylor para várias variáveis, $g \equiv 0$, absurdo. Logo, os zeros de g formam uma superfície algébrica, de modo que $\mathcal{U}$ é aberto e denso e, consequentemente, $\mathcal{D}$ é denso. □

Teorema 9.4. *O conjunto das matrizes invertíveis com entradas em $\mathbb{C}$ é denso.*

Demonstração. Assim como na demonstração do Teorema 9.3, a ideia é encontrar um subconjunto do conjunto $\mathcal{I}$ das matrizes invertíveis que seja denso. Aqui, consideramos o conjunto

$$\mathcal{V} = \{T \in \mathrm{gl}(n, \mathbb{C}) : \det T \neq 0\} \subset \mathcal{I}.$$

Usando um argumento análogo mostramos que $\det T$ é um polinômio nas entradas de T, de modo que o subconjunto acima é denso e, portanto, $\mathcal{I}$ é denso. □

Esses dois resultados podem parecer surpreendentes, mas na verdade são até intuitivos: se um operador tem autovalor nulo, basta uma pequena perturbação para que ele deixe de ser nulo (diagonalizáveis) e, analogamente, se uma matriz tem dois autovalores iguais, basta uma pequena perturbação para diferenciá-los (invertíveis). Com o Teorema 9.3, podemos dar a desejada demonstração alternativa do Teorema de Cayley-Hamilton, como segue.

Teorema 9.5. *Se T é um operador linear em um espaço vetorial de dimensão finita V e p_T é o polinômio característico de T, então $p_T(T) = \mathbf{0}$.*

Observação. *Note que aqui não podemos simplesmente substituir λ por T no polinômio característico, pois isso nos daria*

$$0 = \det(T - T) = \det(TI - T) = p_T(T),$$

absurdo pois, do lado esquerdo, temos um escalar e, do lado direito, uma matriz de ordem n.

Demonstração. Note que p_T é contínuo em T, já que $p_T(T)$ é uma matriz cujas entradas são polinômios nas entradas de T. Pela continuidade de p_T, para verificar $p_T(T) = \mathbf{0}$ basta verificarmos para as diagonalizáveis. Agora, note que

$$p_{U^{-1}TU} = \det(\lambda I - U^{-1}TU) = \det\left(U^{-1}(\lambda I - T)U\right) = p_T,$$

ou seja, o polinômio característico é invariante por conjugação. Portanto, verificar para as diagonalizáveis é o mesmo que verificar para as diagonais. Tomando $T = \operatorname{diag}(\lambda_i)$, $1 \leq i \leq n$, segue que

$$p_T(T) = \begin{pmatrix} p_T(\lambda_1) & 0 & \cdots & 0 \\ 0 & p_T(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & p_T(\lambda_n) \end{pmatrix} = \mathbf{0}.$$

□

Observação. *Como o teorema vale em $gl(n, \mathbb{C})$, ele também vale em $gl(n, \mathbb{R})$ e $gl(n, \mathbb{Q})$.*

Agora, nosso objetivo será dar uma prova alternativa do Teorema 9.2, extraída da Seção 4 de [1]. Para tal, discorreremos um pouco sobre a demonstração anterior e apresentaremos algumas definições.

O argumento utilizando a ordem lexicográfica (ordem lex) é simples e elegante. Essa ordem é uma ordem total no conjunto dos monômios, determinando um único termo líder em cada polinômio, e isso é, à primeira vista, parte do motivo pelo qual a demonstração funciona: ela nos traz à mente a imagem dos termos de f totalmente ordenados e então retirados um por um, da esquerda para a direita, pela nossa escolha adequada de $g_1, \ldots, g_k$.

Mas, na verdade, os termos não são retirados um por um, pois f e $g_1, \ldots, g_k$ são simétricos. Formar $f - g_1$ não só cancela o termo líder $c_1 x_1^{i_1} \cdots x_n^{i_n}$, mas também todos os seus conjugados, e.g., $c_1 x_1^{i_n} x_2^{i_{n-1}} \cdots x_n^{i_1}$. Vemos, então, que de algum modo a ordem lex esconde a simetria entre os conjugados quando coloca um deles como termo líder, ainda que explore a simetria para fazer a prova funcionar.

De fato, a ordem na qual o algoritmo dado na demonstração opera nos termos de f não é realmente a ordem lex. Na verdade, é a ordem que a ordem lex induz no conjunto

das *classes de conjugação* dos termos. A primeira dessas classes é a que contém o termo líder lexicográfico, a segunda contém o maior termo lexicográfico não contido na primeira, e assim por diante. Podemos chamar essa ordem de *ordem lexicográfica simétrica* (ou ordem lex simétrica). Note que ela não é mais uma ordem total nos monômios, apenas nas classes de conjugação. Portanto, o uso da ordem lex na demonstração é enganoso, pois a ordem real é outra.

Isso suscita a pergunta: há descrições da ordem lex simétrica que não passam pela ordem lex? Para respondê-la, precisamos da seguinte definição.

Definição 9.3. *Dado um monômio $x_1^{i_1} \cdots x_n^{i_n}$, defina sua dispersão como $i_1^2 + \cdots + i_n^2$.*

Em termos estatísticos, isso é equivalente (no sentido de que induz a mesma ordem) à variância do conjunto dos expoentes. A dispersão também é equivalente à altura do centro de gravidade do monômio pensado como uma pilha de tijolos, com uma pilha de i_k tijolos correspondente para cada x_k (mostraremos isso a seguir). Além disso, a dispersão é um inteiro não negativo, o que nos permite utilizá-la como base de indução.

O fato chave a ser mostrado é que assim como $c_1 x_1^{i_1} \cdots x_n^{i_n}$, com $i_1 \geq i_2 \geq \cdots \geq i_n$, é o termo líder de $c_1 \sigma_1^{i_1 - i_2} \sigma_2^{i_2 - i_3} \cdots \sigma_n^{i_n}$ na ordem lex, ele e todos seus conjugados também têm dispersão estritamente maior do que os demais termos desse último produto.

Lema 9.6 (Dispersão)**.** *Dados $i_1, \ldots, i_n$ com $i_1 \geq i_2 \geq \cdots \geq i_n$, os termos de*

$$\sigma_1^{i_1 - i_2} \sigma_2^{i_2 - i_3} \cdots \sigma_n^{i_n}$$

com dispersão máxima são $x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}$ e seus conjugados.

Demonstração. A demonstração é bastante interessante, pois se utiliza de uma visualização dos polinômios como pilhas de tijolos. Mais precisamente, identificamos o monômio $x_1^{j_1} \cdots x_n^{j_n}$ com uma sequência de pilhas de alturas $j_1, \ldots, j_n$ de tijolos idênticos. Observe a figura.

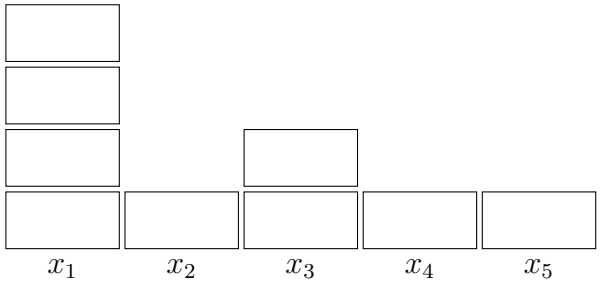


Figura 9: O monômio $x_1^4 x_2 x_3^2 x_4 x_5$

Primeiro, argumentamos que para termos de $\sigma_1^{i_1 - i_2} \sigma_2^{i_2 - i_3} \cdots \sigma_n^{i_n}$, a dispersão é uma função linear crescente da coordenada vertical y do centro de gravidade da configuração em blocos correspondente.

De fato, supondo que cada bloco tem massa unitária, a coordenada vertical do centro de gravidade é dada pela soma, sobre os blocos, das alturas de cada bloco, dividida pela quantidade de blocos. Supondo que o primeiro bloco de cada pilha está à altura 1 e cada bloco tem altura unitária, então a pilha de altura j_1 contribui

$$1 + 2 + \cdots + j_1 = \frac{j_1(j_1 + 1)}{2}$$

para a soma. A coordenada vertical y do centro de gravidade é então dada por

$$\begin{aligned} y &= \frac{1}{d}\left(\frac{j_1(j_1+1)}{2} + \cdots + \frac{j_n(j_n+1)}{2}\right) \\ &= \frac{1}{2d}(j_1^2 + \cdots + j_n^2 + j_1 + \cdots + j_n) \\ &= \frac{1}{2d}(s+d), \end{aligned}$$

sendo d a quantidade de blocos, i.e., o grau do monômio, e s sua dispersão. Logo, $s = 2dy - d$ e, como d é fixo, s é função linear crescente de y.

Agora, observe que todos os termos do produto $\sigma_1^{i_1-i_2}\sigma_2^{i_2-i_3}\cdots\sigma_n^{i_n}$ podem ser obtidos de $x_1^{i_1}\cdots x_n^{i_n}$ movendo blocos horizontalmente (e soltando-os no topo da pilha abaixo se necessário. Os conjugados de $x_1^{i_1}\cdots x_n^{i_n}$ são os termos para os quais cada camada de blocos repousa completamente sobre a camada inferior antes que os blocos sejam soltos. Portanto, os blocos cairão justamente para os termos que não são conjugados de $x_1^{i_1}\cdots x_n^{i_n}$. Observe a figura.

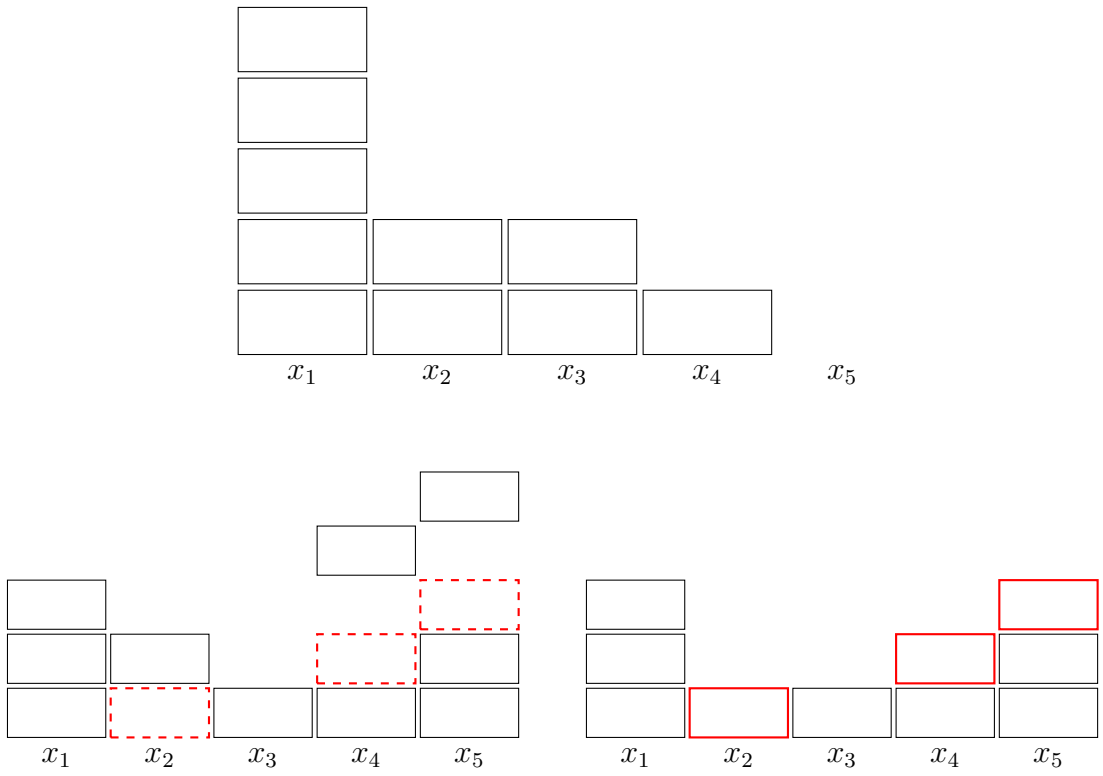


Figura 10: Topo: termo alvo $x_1^5x_2^2x_3^2x_4$. Abaixo: o termo $x_1^3x_2x_3x_4^2x_5^3$ de $\sigma_1^3\sigma_3\sigma_4$ à esquerda e, à direita, o mesmo termo genérico com os blocos "soltos"; ele tem centro de gravidade mais baixo que o termo alvo.

Finalmente, observamos o simples fato de que dada qualquer configuração física de blocos, mover alguns para posições mais baixas diminui a altura do centro de gravidade. □

Com esse lema estabelecido, a demonstração do teorema fundamental segue a mesma linha do argumento padrão dado na demonstração anterior.

Prova do TFFS usando o Lema da Dispersão. Seja f a função simétrica a ser representada. Como antes, suponhamos s.p.g. que f é homogênea. O algoritmo procede da mesma maneira, mas com a dispersão substituindo a ordem lex. Escolha qualquer termo de f com dispersão máxima s_1, e considere ele e seus conjugados. Forme o produto de funções simétricas elementares g_1 que tem esses termos como seus termos de dispersão máxima (se os termos de f têm coeficiente c_1 e expoentes $i_1 \geq i_2 \geq \cdots \geq i_n$, então $g_1 = \sigma_1^{i_1-i_2}\sigma_2^{i_2-i_3}\cdots\sigma_n^{i_n}$). Pelo Lema 9.6, esses termos são os únicos termos de g_1 com dispersão s_1. Daí, $f - g_1$ contém menos termos com dispersão s_1 do que f, possivelmente zero.

Continuando desse modo começando com $f - g_1$, formando g_2 e $f - g_1 - g_2$, e assim por diante, temos um algoritmo que eventualmente termina pois, em cada etapa, ou a dispersão máxima ou a quantidade de termos com tal dispersão diminui.

A unicidade da representação segue como antes: produtos distintos de funções elementares simétricas terão termos distintos com dispersão máxima pela injetividade da correspondência

$$(k_1, \ldots, k_n) \mapsto (k_1 + \cdots + k_n, \ldots, k_n).$$

Logo, é impossível haver um cancelamento total: toda função não nula nas funções elementares simétricas serão não nula quando multiplicada. □

Referências

[1] Blum-Smith, B., Coskey, S., *The Fundamental Theorem on Symmetric Polynomials: History's First Whiff of Galois Theory*, arXiv:1301.7116v5, 2020.

[2] Hoffman, K., Kunze, R., *Linear Algebra*, Second Edition, Prentice-Hall, 1971.