RODOLFO SILVA ALVES

RAÍZES DE POLINÔMIOS: DE BHASKARA A ABEL

Santo André, 2016

UNIVERSIDADE FEDERAL DO ABC

CENTRO DE MATEMÁTICA, COMPUTAÇÃO E COGNIÇÃO

RODOLFO SILVA ALVES

RAÍZES DE POLINÔMIOS: DE BHASKARA A ABEL

Orientador: Prof. Dr. Sinuê Dayan Barbero Lodovici

Dissertação de mestrado apresentada ao Centro de Matemática, Computação e Cognição para obtenção do título de Mestre

ESTE EXEMPLAR CORRESPONDE A VERSÃO FINAL DA DISSERTAÇÃO
DEFENDIDA PELO ALUNO RODOLFO SILVA ALVES,
E ORIENTADA PELO PROF. DR. SINUÊ DAYAN BARBERO LODOVICI.

SANTO ANDRÉ, 2016

Sistema de Bibliotecas da Universidade Federal do ABC
Elaborada pelo Sistema de Geração de Ficha Catalográfica da UFABC
com os dados fornecidos pelo(a) autor(a).

Silva Alves, Rodolfo
Raízes de Polinômios : de Bhaskara a Abel / Rodolfo Silva Alves. — 2016.

165 fls. : il.

Orientador: Prof. Dr. Sinuê Dayan Barbero Lodovici

Dissertação (Mestrado) — Universidade Federal do ABC, Mestrado Profissional em Matemática em Rede Nacional - PROFMAT, Santo André, 2016.

1. Resolução de Equações. 2. Grupos, Anéis e Corpos. 3. Polinômios. 4. Teoria de Galois. 5. Teorema de Abel. I. Barbero Lodovici, Prof. Dr. Sinuê Dayan. II. Mestrado Profissional em Matemática em Rede Nacional - PROFMAT, 2016. III. Título.

MINISTÉRIO DA EDUCAÇÃO
Fundação Universidade Federal do ABC
Programa de Pós-Graduação em Mestrado Profissional em Matemática em Rede Nacional
Avenida dos Estados, 5001 – Bairro Santa Terezinha – Santo André – SP
CEP 09210-580 · Fone: (11) 4996-0017
profmat@ufabc.edu.br

FOLHA DE ASSINATURAS

Assinaturas dos membros da Banca Examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado do candidato Rodolfo Silva Alves, realizada em 19 de agosto de 2016:

Prof.(a) Dr.(a) **Sinuê Dayan Barbero Lodovici** (UFABC) – Presidente

Prof.(a) Dr.(a) **Marcus Antônio Mendonça Marrocos** (UFABC) – Membro Titular

Prof.(a) Dr.(a) **Gleiciane da Silva Aragão** (UNIFESP) – Membro Titular

Prof.(a) Dr.(a) **Daniel Miranda Machado** (UFABC) – Membro Suplente

Prof.(a) Dr.(a) **Renato Alessandro Martins** (UNIFESP) – Membro Suplente

Este exemplar foi revisado e alterado em relação à versão original, de acordo com as observações levantadas pela banca no dia da defesa, sob responsabilidade única do autor e com a anuência de seu orientador.

Santo André, 04 **de** novembro **de 20**16**.**

Assinatura do autor: _____

Assinatura do orientador: _____

Dedico este trabalho a minha esposa e a meus pais

AGRADECIMENTOS

Agradeço primeiro a Deus. Também quero agradecer algumas pessoas que ao longo da minha vida, de certo modo me possibilitaram trilhar este caminho. Assim, agradeço a minha professora Nair da terceira série do antigo primário, ela foi a primeira a me mostrar a beleza da matemática. Ao meu professor da graduação Herivelto Martins Borges Filho, que com muito desprendimento se reunia comigo e mais um amigo, a fim de estudarmos matemática por horas, sem ser para algum exame ou algo do gênero e sim pelo gosto de ensinar matemática.

Também agradeço ao meu orientador Sinuê Dayan Barbero Lodovici, que sempre se mostrou muito dedicado a este trabalho e atento aos detalhes. Mostrando caminhos e possibilidades e não deixando dúvidas para trás.

Agradeço aos meus pais, meu pai Edvaldo Alves, que eu acreditava saber tudo quando era criança. Ele me ajudou nas primeiras lições, que para mim pareciam impossíveis resolver. Herdei dele alguns dos meus primeiros livros universitários. A minha mãe Maria Lúcia da Silva Alves, que desde criança me ensinou o valor da educação, zelando pelos meus estudos sendo sempre presente e dedicada.

E finalmente agradeço a minha esposa Cristiane Aparecida da Silva Alves, que me ajudou nesta caminhada até aqui, sempre com uma palavra de carinho nos momentos mais difíceis. Incentivando-me e encorajando-me neste trajeto tendo sempre muita dedicação e paciência comigo.

A todos vocês o meu muito obrigado!

"Um especialista em resolver problemas deve ser dotado de duas qualidades incompatíveis - uma imaginação inquieta e uma paciente obstinação."

(Howard W. Eves, *matemático estadunidense, 10 de janeiro de 1911 - 6 de junho de 2004*)

RESUMO

O objetivo desta dissertação é estudar as estruturas básicas da álgebra e também os conceitos algébricos relacionados a teoria de Galois, para assim, demonstrar o teorema de Abel. Além disso, mostraremos como resolver as equações quadráticas, cúbicas e quárticas por radicais. Finalizaremos, argumentando sobre o método de Newton, pelo qual podemos resolver numericamente equações que não são solúveis por radicais.

Palavras-chave: Grupos, Anéis, Corpos, Resolução por Radicais, Espaço Vetorial, Método de Newton.

ABSTRACT

The main objective of this thesis is to present basic structures of algebra and some algebraic concepts related to Galois theory, in order to prove Abel's theorem. In addition, we'll show you how to solve quadratic equations, cubic and quartic equations by radicals. Finally, we present Newton's method, by which we can numerically solve equations that are not solvable by radicals.

Keywords: Groups, Rings, Fields, Resolution by Radicals, Vector Space, Newton Method.

CONTEÚDO

INTRODUÇÃO

As equações que estudamos na escola de maneira bem ordenada, podem nos sugerir que o avanço nos estudos destas equações tenham ocorrido de maneira gradual e que foi de modo contínuo. O currículo escolar é estruturado para que o aprendizado ocorra de modo gradual e que sejam *"fornecidas peças"* de tal sorte que sempre consigamos resolver os problemas que nos são dados. Porém, a história da matemática nos mostra que seu desenvolvimento é muito rico e possui muitas peculiaridades. Muitos livros didáticos destacam passagens da história da matemática como curiosidades (incluindo o que segue no parágrafo seguinte).

O Capítulo 1 é dedicado às equações quadráticas, que tem uma fórmula para sua resolução e no Brasil é conhecida como *fórmula de Bhaskara*(1114-c.1185). Curiosamente a fórmula apesar de ter seu nome não foi desenvolvida por Bhaskara (*é sabido que em sua época já era conhecido o método para sua resolução* [12]). Contudo o indiano Bhaskara foi o mais importante matemático do século XII [2]. Foi ele quem preencheu algumas lacunas na obra de Brahmagupta(598-668), por exemplo considerando o problema da divisão por zero. Com efeito, no livro *Vija-Ganita* de Bhaskara, achamos pela primeira vez a afirmação de que um tal quociente é infinito. Outro matemático, o persa ***Mohammed ibu-Musa al-Khwarizmi*** (c.780-c.850) teve importante papel na história da matemática. Um dos livros árabes mais importantes da Idade Média, é um tratado árabe sobre equações chamado de ***al-jabr w'al mûqabala***, cuja tradução pode ser: "*Ciência da reintegração e equiparação*", é de sua autoria. O termo "*álgebra*" tem sua origem neste livro. A palavra ***al-jabr***, ou "*álgebra*", em árabe, era utilizada para designar "restauração", uma das operações usadas na resolução de equações [12].

Problemas que envolvem equações quadráticas e até cúbicas, são bem antigos. Os povos babilônicos, já lidavam com problemas que envolviam equações quadráticas há 4000 a.C. e no caso das equações cúbicas de aproximadamente 2000 à 1600 a.C.. Estas equações cúbicas, que trataremos no Capítulo 2, tiveram a primeira fórmula para sua resolução somente na renascença, envolvendo vários personagens com uma trama bem conturbada. Com este avanço na resolução de equações cúbicas, em meio as fórmulas para as suas resoluções é que nasceu a motivação que possibilitou o surgimento dos

números complexos. Um destes personagens, foi Cardano que publicou em 1545 a *Ars Magna* contendo métodos para resolver equações cúbicas e quárticas. A forma como era resolvida uma equação de quarto grau será descrita no Capítulo 3.

Podemos dizer que estes capítulos, em certo tamanho, encerram a primeira parte da dissertação. A segunda parte, tem início no Capítulo 4, apresentando as principais ferramentas da álgebra, que utilizaremos para mostrar que uma equação de quinto grau não possui uma fórmula para sua resolução, como nos casos estudados anteriormente. Assim trataremos de algumas estruturas algébricas como grupo, anel e corpo.

O Capítulo 5, trata da álgebra linear, um assunto geralmente conhecido por estudantes de matemática. É importante para o nosso objetivo, a utilização de vários dos seus recursos.

Finalmente chegamos ao Capítulo 6, o coração da dissertação. Nele chegamos ao resultado pretendido, mostrar o teorema de Abel e concluir que não é possível encontrar as raízes de uma equação de quinto grau por meio de uma fórmula por radicais. Para isto, ainda iremos percorrer um bom caminho, até que consigamos entender o que diz o teorema de Abel. Ou seja, falaremos um pouco sobre a teoria de Galois.

A tarefa de mostrar que uma equação de quinto grau ou superior não é resolúvel por radicais teve como problema inicial a busca por referências, principalmente no idioma português. Vencida esta etapa, a principal referência para os capítulos 4 e 6 foi a obra [6] que está traduzida para o português. Isto, foi um estímulo a mais para que estes capítulos pudessem suprir algumas dificuldades que eu encontrei na busca da solução do meu propósito. Como o público alvo deste trabalho difere da principal obra de referência, tive cautela com a linguagem, adequando quando foi necessário. Por ser tratar de um tema sensível, tomei o cuidado de incluir alguns exemplos e demonstrar os teoremas completamente, uma vez que no texto original há algumas omissões nas passagens de alguns destes teoremas. Também resolvi alguns dos exercícios propostos pelo autor, para uma melhor compreensão de um ou outro teorema. O leitor pode encontrar mais em [1] e [4], ambos em inglês. O capítulo 5 que também possui papel importante para construção da estratégia tomada para a demonstração do teorema de Abel, têm como referências [6] e [3]. Neste capítulo, mesclei as referências citadas e as ordenei de modo a auxiliar o entendimento do capítulo 6.

Encerrando o trabalho, o Capítulo 7 é dedicado ao método de Newton que nos possibilita determinar não somente as equações de quinto grau, como muitas outras de maneira aproximada, mas com uma precisão tão boa quanto se queira. Este método

é bom até mesmo para calcular as equações cúbicas, pois as fórmulas de Cardano são bem difíceis de se usar.

1

EQUAÇÕES QUADRÁTICAS

Os problemas que podem ser resolvidos através de resolução de uma equação quadrática são muito antigos, alguns datam há quase quatro mil anos, escritos pelos babilônios. Bhaskara, um nome bem conhecido pela fórmula que não é sua, também curiosamente pode ser confundido por outro, pois existem dois, um nascido em 629 e outro em 1114 [12]. Assim são também conhecidos respectivamente de Bhaskara I e II. O Bhaskara I escrevia seus comentários nos livros indianos de matemática que eram escritos em versos e tinham difícil compreensão, e por isso era comum a prática de outros matemáticos escreverem comentários para auxiliar o entendimento aos leitores. O comentário mais antigo de Bhaskara I, foi sobre o livro de Aryabhata. Contudo o personagem principal desta história é Bhaskara II, autor dos livros mais populares de aritmética e álgebra no século XII, que citou os procedimentos utilizados por Brahmagupta (Bhinmal, Rajastão, 598-668). No Apêndice A.1 está descrito de que forma Bhaskara II resolvia as equações quadráticas.

Na época de Bhaskara não se usavam fórmulas, as equações eram descritas por retóricas o que mudou com François Viète, matemático francês que viveu de 1540 a 1603, nesta época foi introduzido o uso de fórmulas com coeficientes o que mudou a maneira de resolver uma equação quadrática. Agora é possível atacar o problema por um outro olhar, utilizando a álgebra, podemos dizer *grosso modo* que a álgebra é uma tradução da retórica usada por Bhaskara, por exemplo, em uma linguagem matemática. No sétimo ano do ensino fundamental, os alunos começam a aprender álgebra, e começam com exercícios como passar a expressão *o dobro de um número* da linguagem corrente, para uma expressão matemática, no caso, $2x$. Nos dias de hoje, é interessante notar que é comum ouvir expressões algébricas, como *"João me disse, que me pagaria um x valor para resolver o seu problema"*, nos mostra que nossa linguagem cotidiana recebeu influência da linguagem matemática. Entretanto, não significa que

é uma linguagem matemática. Não há garantias de quem falou tenha conhecimento de álgebra, no caso x é simplesmente sinônimo de determinado.

1.1 CONJUNTOS E FUNÇÕES

Antes de partirmos para resolução de uma equação quadrática será necessário algumas informações preliminares. Para resolver uma equação ou outros problemas a Matemática atualmente se utiliza da linguagem de conjuntos. A noção básica de conjuntos não é definida, ou seja, é aceita intuitivamente e, por isso, chamada *noção primitiva*. Ela foi utilizada primeiramente por Georg Cantor (1845-1918), matemático nascido em São Petersburgo, mas que passou a maior parte de sua vida na Alemanha. Segundo Cantor, a noção de conjunto designa uma coleção de objetos bem definidos e discerníveis, chamados elementos do conjunto.

Os conjuntos numéricos são de grande importância na matemática, o primeiro conjunto que aprendemos é o conjunto dos números naturais $\mathbb{N}$ que usamos principalmente para a contagem. Temos:

$$\mathbb{N} = \{1, 2, 3, \ldots\}.$$

Alguns autores consideram o zero como um número natural. A opção por uma ou outra alternativa é uma questão de gosto ou de conveniência. Entretanto, como citado em [11] *"o zero foi empregado inicialmente pelos maias, posteriormente pelos hindus, difundido pelos árabes e adotado no ocidente, não como um número e sim como um algarismo, com o utilíssimo objetivo de preencher uma casa decimal vazia"*.

Será também de grande valia utilizarmos o conceito de *função* [10].

Definição 1.1. *Função*
Dados os conjuntos X e Y, uma função $f : X \to Y$ (leia "uma função de X em Y") é uma regra (ou conjunto de instruções) que diz como associar a cada elemento $x \in X$ um elemento $y = f(x) \in Y$ (leia "y igual a f de x"). O conjunto X chama-se *domínio* e Y chama-se *imagem* de x pela função f, ou *valor* assumido pela função f no ponto $x \in X$. Escreve-se $x \mapsto f(x)$ para indicar que f transforma (ou leva) x em $f(x)$.

Definição 1.2. *Gráfico*
O gráfico de uma função $f : X \to Y$ é o subconjunto $G(f)$ do produto cartesiano $X \times Y$ formado por todos os pares ordenados (x, y), onde x é um ponto qualquer de X e $y = f(x)$. Assim,

$$G(f) = \{(x, y) \in X \times Y; y = f(x)\} = \{(x, f(x)) ; x \in X\}.$$

Alguns exemplos de funções:

Exemplo 1.1. *Função Identidade*
Seja X um conjunto qualquer, não vazio. Chamamos de *função identidade* a função $f : X \to X$, tal que $f(x) = x$ para todo $x \in X$.

Exemplo 1.2. *Função Constante*
Sendo X e Y dois conjuntos quaisquer, não vazios. Chamamos de *função constante* a função $f : X \to Y$, tal que $f(x) = c$ para todo $x \in X$, onde $c \in Y$.

Uma função $f : X \to Y$ é chamada *injetiva* quando elementos diferentes em X são transformados por f em elementos diferentes em Y. Ou seja, f é injetiva quando

$$x \neq x' \text{ em } X \Longrightarrow f(x) \neq f(x').$$

Diremos que uma função $f : X \to Y$ é chamada *sobrejetiva* quando para qualquer elemento $y \in Y$, temos pelo menos um elemento $x \in X$ tal que $f(x) = y$.

Uma função $f : X \to Y$ chama-se uma *bijeção*, ou uma *correspondência biunívoca* entre X e Y quando é ao mesmo tempo injetiva e sobrejetiva.

Agora podemos apresentar uma definição de contagem de conjunto. Assim vamos definir o *número cardinal* de um conjunto e o que são conjuntos finitos e infinitos.

Definição 1.3. Contar um conjunto X significa estabelecer uma correspondência biunívoca entre os elementos de X e os de um subconjunto de $\mathbb{N}$ da forma $I_n = \{x \in \mathbb{N}; x \leqslant n\} = \{1, 2, \ldots, n\}$. Quando é possível estabelecer tal correspondência biunívoca, dizemos que X é um conjunto finito e que n é o *número cardinal* ou *número de elementos* de X.

Um conjunto é infinito quando não é finito. Apesar de parecer num primeiro momento, que todos os conjuntos infinitos têm o mesmo "tamanho", isso não ocorre. Georg Cantor mostrou isso ao comparar a cardinalidade de conjuntos infinitos.

Definição 1.4. Dizemos que dois conjuntos X e Y têm a mesma cardinalidade quando é possível estabelecer uma correspondência biunívoca entre X e Y (isto é, existe uma função bijetiva $f : X \to Y$).

Outros dois conjuntos, que também têm destaque entre os conjuntos numéricos, o conjunto dos números inteiros $\mathbb{Z}$ e o conjunto dos números racionais $\mathbb{Q}$. O conjunto dos números inteiros é obtido adicionando os números negativos aos números naturais (neste caso estou considerando o zero como sendo um número natural). O conjunto dos números racionais é formado pela razão entre dois inteiros, onde o denominador é diferente de zero. Assim temos:

$$\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$$

e

$$\mathbb{Q} = \left\{ \frac{x}{y};\ x \in \mathbb{Z} \text{ e } y \in \mathbb{Z}\backslash\{0\} \right\}.$$

Podemos estabelecer uma correspondência biunívoca dois a dois, entre os conjuntos $\mathbb{N}$, $\mathbb{Z}$ e $\mathbb{Q}$. Assim, estes conjuntos têm a mesma cardinalidade. Dizemos que um conjunto é *enumerável* quando o conjunto é finito ou se o conjunto tiver a mesma cardinalidade de $\mathbb{N}$. Quando um conjunto não é enumerável é chamado de *não enumerável*. Portanto, $\mathbb{N}$, $\mathbb{Z}$ e $\mathbb{Q}$ são conjuntos enumeráveis.

Os últimos dois conjuntos numéricos que iremos comentar neste capítulo são o dos números reais $\mathbb{R}$ e dos números irracionais. O conjunto dos números reais pode ser visto como o modelo aritmético de uma reta. Tomando uma reta r e marcando uma medida arbitrária m podemos fazer uma correspondência entre os pontos tomados a partir de um ponto O qualquer. Assim se fizermos o ponto O corresponder ao 0, temos a cada medida m a direita, a partir de O corresponder aos números naturais , se tomarmos medidas m a esquerda, teremos os números negativos, e O corresponderá ao número 0.

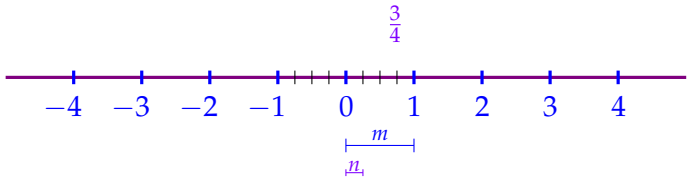

Figura 1: Reta real.

É possível determinar a posição de um número racional dividindo a unidade de comprimento m em n partes iguais e tomando quantas partes n for necessário para o número racional escolhido. Por exemplo, para encontrarmos $3/4$ na reta real, dividimos m em 4 partes iguais que indicarei por n, assim se tomarmos $3n$ a partir de O, obtemos o número $3/4$ conforme indicado na Figura 1. Uma descrição mais completa é encontrada em [10].

Os números que podemos encontrar na reta com o auxílio de segmentos *comensuráveis*, são números racionais. Nem todos os números podem ser encontrados desta maneira, ou seja, partindo do segmento unitário não é possível dividi-lo em segmentos iguais de modo que possamos encontrar o número por intermédio destes segmentos, assim chamamos estes segmentos de *incomensuráveis*. Os números que não podem ser obtidos deste modo são chamados de números irracionais e completam a reta real. Livros de ensino médio, como por exemplo [7], usam $\mathbb{I}$ para denotar o conjuntos dos números irracionais. A união dos números racionais com os números irracionais resulta no conjunto dos números reais que é denotado por $\mathbb{R}$. O conjunto dos números reais e irracionais são exemplos de conjuntos não-enumeráveis. O leitor pode encontrar a descrição do conjunto dos números reais de maneira mais precisa em [9].

Finalizamos esta seção com a definição de função quadrática.

Definição 1.5. *Função Quadrática*
É uma função $f : \mathbb{R} \to \mathbb{R}$ onde existem números reais a,b, c, com $a \neq 0$, tais que $f(x) = ax^2 + bx + c$ para todo $x \in \mathbb{R}$.

1.2 FATORAÇÃO: CASO DO TRINÔMIO DE SEGUNDO GRAU

No antigo ginásio, hoje fundamental II, estudamos vários casos de fatoração. Um caso particular de fatoração é usada para resolver algumas equações quadráticas. Em [7] o clássico trinômio $x^2 - 5x + 6$ é fatorado em $(x - 2)(x - 3)$. Neste exemplo, o

autor mostra aos alunos que se o polinômio têm raízes x_1 e x_2, então $x - x_1 = 0$ e $x - x_2 = 0$. Assim a forma fatorada do trinômio do 2° grau em x, $x^2 - (x_1 + x_2)x + x_1x_2$, é $(x - x_1)(x - x_2)$. O autor também destaca que, dadas as raízes x_1 e x_2, $x_1 + x_2$ é a soma das raízes e que x_1x_2 é seu produto. Deste modo temos um método para resolver uma equação quadrática do tipo $x^2 - sx + p = 0$, onde s é a soma das raízes e p é o produto delas.

Como dissemos no início do capítulo, os problemas que envolvem equações quadráticas são muito antigos, alguns datam há quase quatro mil anos, escritos pelos babilônios. Entre estes textos, em escrita cuneiforme [1] que os babilônios trouxeram, encontram-se questões que procuravam dois números conhecendo sua soma s e seu produto p.

Geometricamente, podemos escrever o problema do seguinte modo:

Problema 1. Determine os lados de um retângulo conhecendo o seu semiperímetro s e a sua área p.

Evidentemente, naquela época não havia o recurso algébrico que dispomos hoje. Como mencionado acima, até o fim do século XVI, não se usava a linguagem algébrica. Entretanto, a usaremos.

Primeiro, vamos mostrar que a solução do Problema 1 apresentado é a solução da equação $x^2 - sx + p = 0$. Para isto, note que fazendo x, um dos números, o outro será $s - x$ e assim:

$$p = x(s - x) = sx - x^2,$$

logo

$$x^2 - sx + p = 0.$$

1 A escrita cuneiforme foi desenvolvida pelos sumérios, sendo a designação geral dada a certos tipos de escritas feitas com auxílio de objetos em formato de cunha. É juntamente com os hieróglifos egípcios, o mais antigo tipo conhecido de escrita, tendo sido criado pelos sumérios por volta de 3500 a.C.. Inicialmente a escrita representava formas do mundo (pictogramas), mas por praticidade as formas foram se tornando mais simples e abstratas.

Afirmamos que se x é uma solução $s - x$ também o é, pois note que se α é uma solução teremos $\beta = s - \alpha$ a outra solução, assim:

$$\begin{aligned}\beta^2 - s\beta + p &= (s-\alpha)^2 - s(s-\alpha) + p \\ &= s^2 - 2s\alpha + \alpha^2 - s^2 + s\alpha + p \\ &= \alpha^2 - s\alpha + p = 0.\end{aligned}$$

Antes de continuarmos, vamos escrever a solução para o Problema 1.

Solução: *Eleve ao quadrado a metade da soma, subtraia o produto e extraia a raiz quadrada da diferença. Some ao resultado a metade da soma. Isso dará o maior dos números procurados. Subtraia-o da soma para obter o outro número.*

Como o problema apresentado é equivalente a resolver a equação $x^2 - sx + p = 0$, podemos escrever a regra que nos fornece as raízes modernamente da seguinte forma:

$$x = \frac{s}{2} + \sqrt{\left(\frac{s}{2}\right)^2 - p} \qquad \text{e} \qquad s - x = \frac{s}{2} - \sqrt{\left(\frac{s}{2}\right)^2 - p}.$$

Não há registro dos autores dos textos cuneiformes dos argumentos que os levaram a tal solução, entretanto existem indícios de que eles poderiam ter usado algum argumento semelhante ao que mostraremos a seguir.

Sejam α e β, com $\alpha \leq \beta$, as soluções para a equação $x^2 - sx + p = 0$. Sabemos que a média aritmética entre α e β é

$$\frac{\alpha + \beta}{2} = \frac{s}{2},$$

assim os números α e β são equidistantes da média aritmética. Desta maneira, basta conhecer a diferença $d = \beta - \left(\frac{s}{2}\right) = \left(\frac{s}{2}\right) - \alpha$ para chegarmos as raízes da questão. Para isto, primeiro faremos $\alpha = \left(\frac{s}{2}\right) - d$ e $\beta = \left(\frac{s}{2}\right) + d$. Com isto, lembrando que $p = \alpha\beta$, temos:

$$p = \alpha\beta = \left(\frac{s}{2} - d\right)\left(\frac{s}{2} + d\right) = \left(\frac{s}{2}\right)^2 - d^2 \iff d^2 = \left(\frac{s}{2}\right)^2 - p.$$

Portanto

$$d = \sqrt{\left(\frac{s}{2}\right)^2 - p},$$

logo, chegamos as raízes:

$$\alpha = \frac{s}{2} - d = \frac{s}{2} - \sqrt{\left(\frac{s}{2}\right)^2 - p}$$

e

$$\beta = \frac{s}{2} + d = \frac{s}{2} + \sqrt{\left(\frac{s}{2}\right)^2 - p}.$$

1.3 FORMA CANÔNICA DO TRINÔMIO

Uma equação polinomial de segundo grau ou equação quadrática é uma equação do tipo $ax^2 + bx + c = 0$, onde existem números reais a, b e c, com $a \neq 0$. Queremos determinar uma solução para esta equação, para quaisquer que sejam os coeficientes reais a, b e c, com $a \neq 0$, utilizando para isso os coeficientes deste trinômio. Assim considere:

$$ax^2 + bx + c = a\left[x^2 + \frac{b}{a}x + \frac{c}{a}\right]. \tag{1.1}$$

A ideia é reescrever o trinômino (1.1) acima de tal modo que consigamos eliminar o termo x^2. Desta forma, lembrando que o Trinômio Quadrado Perfeito é:

$$(X + Y)^2 = X^2 + 2XY + Y^2,$$

onde X e Y são números reais. Deste modo, temos que reescrever (1.1) de modo que obtenhamos um Trinômio Quadrado Perfeito. Logo temos:

$$X = x \tag{1.2}$$

$$2XY = \frac{b}{a}x. \tag{1.3}$$

Substituindo (1.2) em (1.3), temos:

$$\begin{aligned} 2XY &= \frac{b}{a}X \\ 2Y &= \frac{b}{a} \\ Y &= \frac{b}{2a}. \end{aligned}$$

Deste modo, podemos escrever o Trinômio Quadrado Perfeito:

$$(X + Y)^2 = \left(x + \frac{b}{2a}\right)^2 = x^2 + 2x\frac{b}{2a} + \left(\frac{b}{2a}\right)^2 = x^2 + \frac{b}{a}x + \frac{b^2}{4a^2}.$$

Retomando a equação (1.1), percebemos que as duas primeiras parcelas dentro do colchete são as mesmas do desenvolvimento do quadrado $\left(x + \frac{b}{2a}\right)^2$, faltando o termo

$\frac{b^2}{4a^2}$ para ser um quadrado perfeito. Desta maneira, vou reescrever a equação (1.1), somando e subtraindo $\frac{b^2}{4a^2}$ de dentro do colchete. Assim obtemos:

$$\begin{aligned} ax^2 + bx + c &= a\left[x^2 + \frac{b}{a}x + \frac{b^2}{4a^2} - \frac{b^2}{4a^2} + \frac{c}{a}\right] \\ &= a\left[\left(x + \frac{b}{2a}\right)^2 + \frac{4ac - b^2}{4a^2}\right]. \end{aligned} \tag{1.4}$$

A maneira de representar o trinômio em (1.4) é chamado de *forma canônica* [10]. Deste modo, podemos determinar em poucos passos a fórmula que nos permite encontrar as raízes da equação quadrática, $ax^2 + bx + c = 0$ com $a \neq 0$. Assim temos as seguintes equivalências:

$$ax^2 + bx + c = 0 \iff \left(x + \frac{b}{2a}\right)^2 + \frac{4ac - b^2}{4a^2} = 0 \tag{1.5}$$

$$\iff \left(x + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a^2} \tag{1.6}$$

$$\iff x + \frac{b}{2a} = \pm\frac{\sqrt{b^2 - 4ac}}{2a} \tag{1.7}$$

$$\iff x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

A passagem da linha (1.6) para linha (1.7) só tem significado quando o *discriminante*

$$\Delta = b^2 - 4ac$$

é maior do que ou igual a 0. No caso de $\Delta < 0$, a equivalência entre as linhas (1.5) e (1.6) significa que a equação dada não possui solução real, pois o quadrado de $x + \frac{b}{2a}$ não pode ser negativo. Também há possibilidade de Δ ser igual à 0, que nos leva a conclusão que $x = -\frac{b}{2a}$. Percebemos que o discriminante (Δ) tem papel importante na fórmula e por isso alguns livros didáticos escrevem a fórmula em duas partes:

$$x = \frac{-b \pm \sqrt{\Delta}}{2a}$$

e

$$\Delta = b^2 - 4ac.$$

E reservam um tópico para a discussão do sinal do Δ.

Resumidamente podemos dizer que se $\Delta < 0$ não há soluções reais. Se a equação quadrática possuir solução real, elas são duas e são chamadas de raízes, que podem ser representadas por x_1 e x_2. Assim as raízes da equação quadrática, são:

$$x_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a}$$

e

$$x_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}.$$

Lembrando que se $\Delta = 0$, temos:

$$x_1 = x_2 = -\frac{b}{2a}.$$

1.4 O GRÁFICO DA FUNÇÃO QUADRÁTICA

O gráfico de uma função quadrática é uma parábola. Denominamos *parábola* $\mathcal{P}$, o *lugar geométrico*[2] formado pelos pontos que equidistam de um ponto dado F e uma reta d que não contém o ponto F. A reta d é denominada diretriz e o ponto F de foco da parábola.

Ainda podemos destacar, além do *foco* F e da *diretriz* d, conforme a Figura 2 os seguintes elementos da parábola:

- A reta perpendicular à diretriz, baixada a partir do foco. Que é denominado *eixo* da parábola;
- O *vértice* V, que é o ponto da parábola mais próximo da diretriz.

Sabendo que a distância de um ponto $P = (x_0, y_0)$ do plano cartesiano a uma reta $r : ax + by = c$ é o comprimento do segmento perpendicular baixado do ponto sobre a reta r, que indicaremos por $d(P, r)$. Este resultado está demonstrado em [3] e pode ser calculado como

$$d(P, r) = \frac{|ax_0 + by_0 - c|}{\sqrt{a^2 + b^2}}.$$

2 Lugar geométrico consiste no conjunto de pontos de um plano que gozam de uma determinada propriedade.

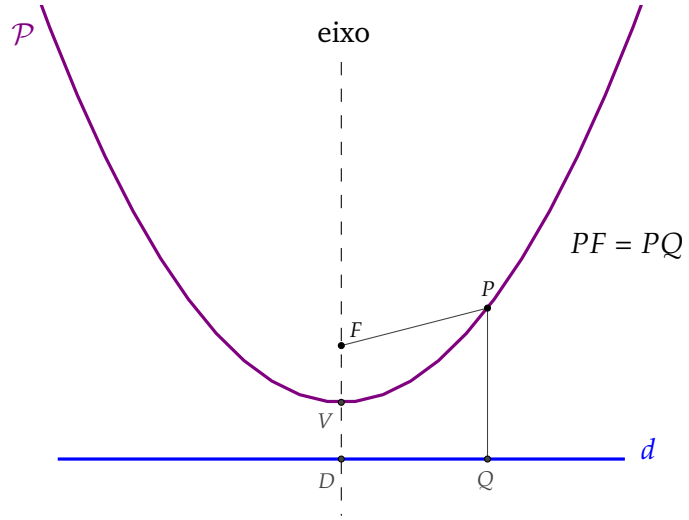

Figura 2: Parábola.

E a distância entre dois pontos $P = (x_1, y_1)$ e $Q = (x_2, y_2)$ no plano cartesiano, que denotaremos por $d(P, Q)$, é a medida da hipotenusa PQ do triângulo retângulo PQR, onde $R = (x_2, y_1)$ e seus catetos são PR e QR (Figura 3). Lembrando que a distância entre dois pontos no mesmo eixo é igual ao módulo da diferença de suas coordenadas, podemos escrever as medidas dos catetos como sendo $d(P, R) = |PR| = |x_2 - x_1|$ e $d(Q, R) = |QR| = |y_2 - y_1|$. Assim, pelo Teorema de Pitágoras, temos

$$d(P, Q) = |PQ| = \sqrt{|PR|^2 + |QR|^2} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

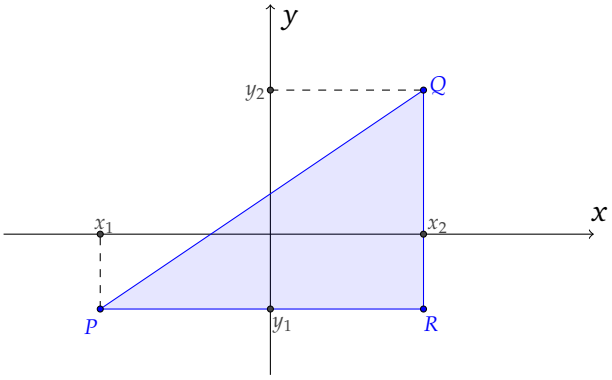

Figura 3: Distância entre dois pontos no plano cartesiano.

Agora, mostraremos no exemplo a seguir que a função $f(x) = x^2$ é uma parábola.

Exemplo 1.3. Suponhamos que o gráfico da função $f(x) = x^2$ é uma parábola, então o vértice V da parábola $\mathcal{P}$ é o ponto médio entre o foco F e a reta diretriz d. Também sabemos que o ponto $(0,0)$ é o vértice da parábola $\mathcal{P}$, pois todo número elevado ao quadrado é um número positivo, assim o menor valor que a função admite é 0. Portanto, considerando D o ponto em que o eixo y, que contém o ponto V, intersecta a reta diretriz d, temos $FV = VD$. Tomando esta distância igual a p, temos $F = (0,p)$ e $D = (0,-p)$, veja a Figura 4. Assim, temos $d : y = -p$. Logo

$$P = (x,y) \in \mathcal{P} \iff \sqrt{x^2 + (y-p)^2} = |y+p| \iff x^2 = 4py \iff y = \frac{x^2}{4p}.$$

Desta forma, se tomarmos $P \in \mathcal{P}$, por exemplo $P = \left(1, 1^2\right) = (1,1)$, temos:

$$P = (1,1) \in \mathcal{P} \iff 1 = \frac{1^2}{4p} \iff p = \frac{1}{4}.$$

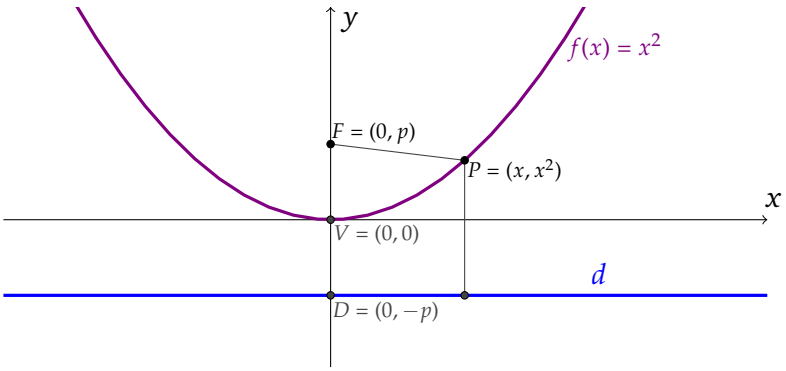

Figura 4: Parábola $y = x^2$.

Portanto a parábola procurada possui $F = \left(0, \frac{1}{4}\right)$ e reta diretriz $d : y = -\frac{1}{4}$. E de fato, esta parábola $\mathcal{P}$ representa o gráfico da função quadrática $f(x) = x^2$, pois para todo $x \in \mathbb{R}$ a igualdade $d(P,F) = d(P,d)$ é válida. Note que:

$$\begin{gathered}
d(P,F) = d(P,d) \\
\sqrt{(x-0)^2 + \left(x^2 - \frac{1}{4}\right)^2} = \frac{\left|0 \cdot x + 1 \cdot x^2 - \left(-\frac{1}{4}\right)\right|}{\sqrt{0^2+1^2}} \\
\sqrt{x^2 + \left(x^2 - \frac{1}{4}\right)^2} = \left|x^2 + \frac{1}{4}\right| \\
x^2 + \left(x^2 - \frac{1}{4}\right)^2 = \left(x^2 + \frac{1}{4}\right)^2 \\
\frac{3x^2}{2} + \frac{1}{16} = \frac{3x^2}{2} + \frac{1}{16}.
\end{gathered}$$

Como vimos na seção anterior, podemos representar qualquer trinômio do tipo $ax^2 + bx + c$ na forma canônica, $a\left[\left(x+\frac{b}{2a}\right)^2 + \frac{4ac-b^2}{4a^2}\right] = a\left(x+\frac{b}{2a}\right)^2 + \frac{4ac-b^2}{4a}$. Esta forma, nos permite manipular o polinômio de uma maneira mais simples, assim tomarei

$$m = \frac{b}{2a} \qquad \text{e} \qquad k = \frac{4ac-b^2}{4a},$$

deste modo, podemos escrever uma função quadrática na forma canônica, como:

$$f(x) = a(x-m)^2 + k.$$

Usando esta forma, exibiremos em casos os gráficos das funções quadráticas. Mostrando assim, que qualquer função quadrática do tipo $f(x) = ax^2 + bx + c$ com $a \neq 0$, possui como gráfico uma parábola.

- Caso: $f(x) = ax^2$ $\qquad (m = 0 \text{ e } k = 0)$

  Exemplo 1.4. Se $a \neq 0$ e $f(x) = ax^2$, procedendo de modo análogo ao Exemplo 1.3 é possível mostrar que o gráfico da função quadrática $f(x) = ax^2$ é a parábola de foco $F = \left(0, \frac{1}{4a}\right)$ e a reta diretriz $d : y = -\frac{1}{4a}$ (Figura 5). E para comprovar este fato, basta mostrar que, para todo $x \in \mathbb{R}$, vale a igualdade $d(P,F) = d(P,d)$, ou seja:

$$x^2 + \left(ax^2 - \frac{1}{4a}\right)^2 = \left(ax^2 + \frac{1}{4a}\right)^2.$$

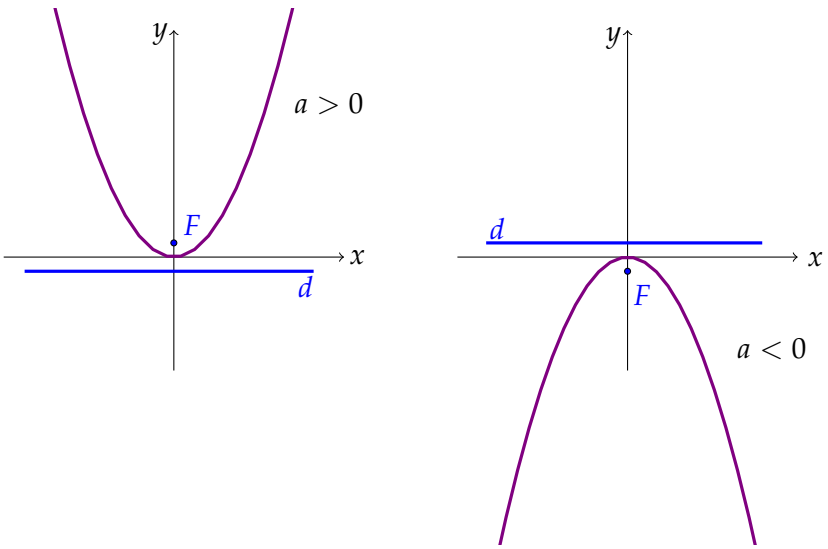

Figura 5: A parábola $f(x) = ax^2$ tem sua concavidade voltada para cima se $a > 0$ e sua concavidade voltada para baixo se $a < 0$.

- Caso: $f(x) = a(x-m)^2$ $\qquad (m \neq 0 \text{ e } k = 0)$

Exemplo 1.5. Neste caso, para todo $a \neq 0$ e todo $m \in \mathbb{R}$, o gráfico da função quadrática $f(x) = a(x-m)^2$ é uma parábola de foco $F = \left(m, \frac{1}{4a}\right)$ e reta diretriz $d : y = -\frac{1}{4a}$. E para comprovar este fato, basta mostrar que, para todo $x \in \mathbb{R}$, vale a igualdade $d(P,F) = d(P,d)$, ou seja:

$$(x-m)^2 + \left[a(x-m)^2 - \frac{1}{4a}\right]^2 = \left[a(x-m)^2 + \frac{1}{4a}\right]^2.$$

Podemos também mostrar o mesmo resultado, observando que o gráfico de $f(x) = a(x-m)^2$ (Figura 6) resulta do gráfico de $g(x) = ax^2$ pela translação horizontal $(x,y) \mapsto (x+m, y)$, a qual leva o eixo $x = 0$ no eixo $x = m$.

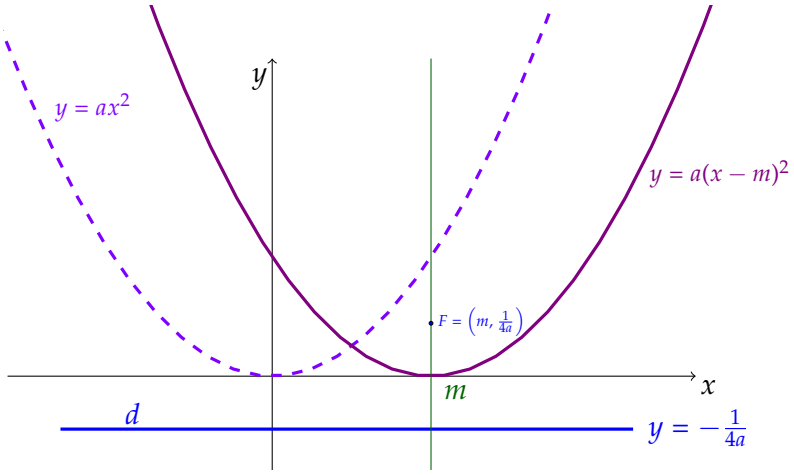

Figura 6: Parábola de foco $F = \left(m, \frac{1}{4a}\right)$ e reta diretriz $d : y = -\frac{1}{4a}$.

▶ Caso: $f(x) = a(x-m)^2 + k \qquad (m \neq 0 \text{ e } k \neq 0)$

Exemplo 1.6. Sejam $a, m, k \in \mathbb{R}$, com $a \neq 0$, o gráfico da função quadrática $f(x) = a(x-m)^2 + k$ é a parábola de foco $F = \left(m, k + \frac{1}{4a}\right)$ e reta diretriz $d : k - \frac{1}{4a}$.

O resultado é imediato a partir do Exemplo 1.5, tomando o gráfico da função quadrática $f(x) = a(x-m)^2 + k$ como resultado de uma translação vertical $(x,y) \mapsto (x, y+k)$ do gráfico de $g(x) = a(x-m)^2$, que leva o eixo das abscissas na reta $y = k$ e a reta $y = -\frac{1}{4a}$ na reta $y = k - \frac{1}{4a}$ (Figura 7).

Pelos exemplos, percebemos que escrever a função quadrática na forma canônica também é vantajosa para a obtenção do gráfico. Do Exemplo 1.6 podemos concluir que o gráfico de qualquer função quadrática

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

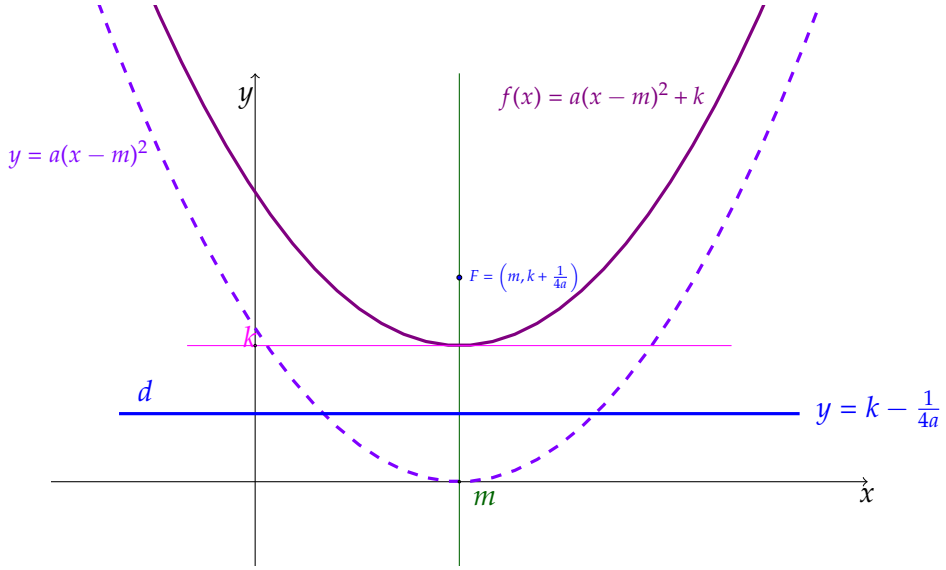

Figura 7: Parábola de foco $F = \left(m, k + \frac{1}{4a}\right)$ e reta diretriz $d : y = k - \frac{1}{4a}$.

é a parábola que possui a reta diretriz

$$d : y = \frac{4ac - b^2 - 1}{4a}$$

e o foco no ponto

$$F = \left(-\frac{b}{2a}, \frac{4ac - b^2 + 1}{4a}\right).$$

Sua concavidade depende do valor de a, se $a > 0$ a concavidade será voltada para cima e se $a < 0$, sua concavidade será voltada para baixo.

Também decorre do exemplo ilustrado que o vértice V do gráfico de uma função quadrática $f(x) = a(x-m)^2 + k$ é $V = (m, k)$, assim o vértice da $f(x) = ax^2 + bx + c$ é

$$V = \left(-\frac{b}{2a}, \frac{4ac - b^2}{4a}\right).$$

Como podemos observar no gráfico, o vértice fornece o maior valor (ou menor) que a função assume, assim a coordenada y do vértice é chamado *valor máximo da função quadrática*, quando $a < 0$ e *valor mínimo da função quadrática*, quando $a > 0$. Assim o vértice é um *ponto máximo*, ou *ponto mínimo* da função quadrática.

Nos livros didáticos é comum fornecer a fórmula do vértice V da parábola utilizando Δ na fórmula. Lembrando que $\Delta = b^2 - 4ac$, temos que $-\Delta = 4ac - b^2$, assim os livros costumam escrever $V = (x_v, y_v)$, com

$$x_v = -\frac{b}{2a} \qquad \text{e} \qquad y_v = -\frac{\Delta}{4a}.$$

Observação 1.6. Não é recomendável dizer que o valor de a altera a boca da parábola, dizendo por exemplo: **"quanto menor o valor de a, mais aberta é a "boca" da parábola"**. Isto não faz sentido, uma vez que pela definição de parábola, todas as parábolas são semelhantes. Do mesmo modo que todos os círculos são semelhantes. Acredito que seria mais prudente dizer que há uma ampliação ou redução da parábola conforme se diminui ou aumenta o valor de a. Observe na Figura 8 as parábolas $f(x) = x^2$ e $g(x) = \frac{1}{8}x^2$.

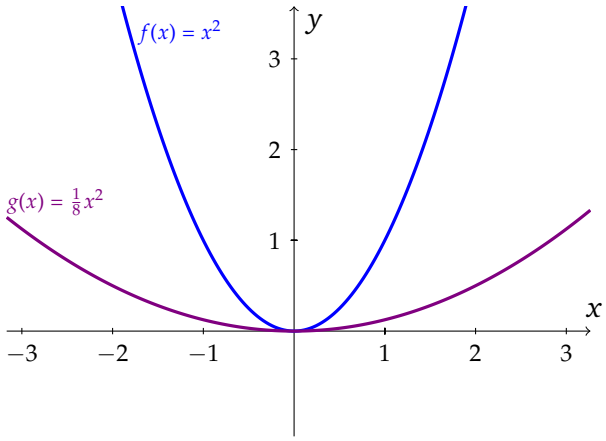

Figura 8: As funções estão representadas num intervalo aproximado de $[-3, 3]$.

Num primeiro momento, pode-se ter uma conclusão precipitada de que $g(x)$ tem a "boca" mais aberta. Porém ao aumentar o domínio em que o gráfico foi desenhado, observamos na Figura 9 que o gráfico da função $g(x)$ é semelhante ao gráfico da função $f(x)$.

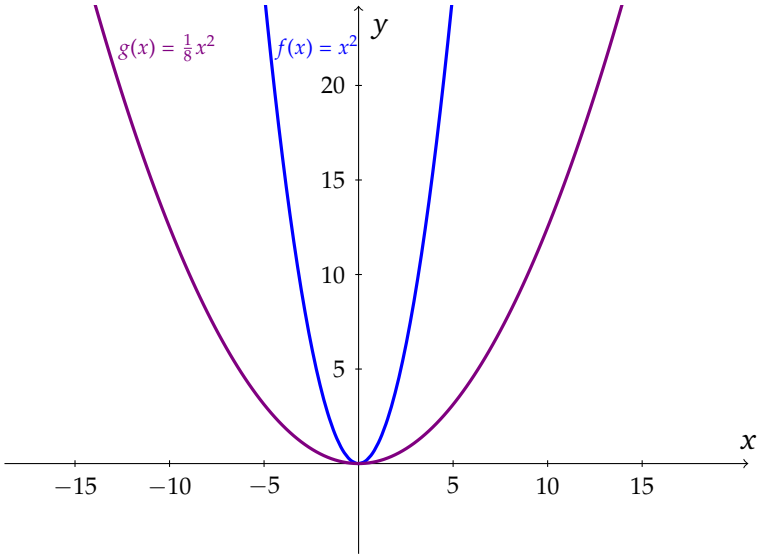

Figura 9: As funções estão representadas num intervalo aproximado de $[-15, 15]$

2

EQUAÇÕES CÚBICAS

Equações cúbicas também eram conhecidas dos antigos babilônicos, além dos gregos, chineses, indianos e egípcios. Os tabletes cuneiformes babilônicos datam de aproximadamente de 2000 à 1600 a.C., lá foram encontradas tabelas para o cálculo de cubos e raízes cúbicas, mas não existem evidências de que por meio destas, os babilônicos tenham resolvido equações cúbicas. Entretanto foi somente na renascença, ou seja, depois de três milênios é que houve um salto na resolução deste problema.

Esta é uma parte um tanto nebulosa da história da matemática. Um dos primeiros avanços para se encontrar uma solução para a equação cúbica foi dada por Scipione del Ferro,[1] que ensinou na Universidade da Bolonha no início do século XVI, foi o primeiro a resolver equações cúbicas, do tipo $x^3 + px^2 = q$, segundo relatou Girolamo Cardano[2], em sua *Ars Magna* [1]. Esta obra publicada em 1545, foi a primeira obra a conter métodos de resolução de equações de terceiro e também de quarto grau.

Sem tornar público o seu método, del Ferro o ensina ao seu aluno Antonio Maria del Fiore, ou, Fior, como também é conhecido. Nesta época, Tartaglia [3], se dedicou a encontrar uma solução para equação cúbica de modo independente e por volta de 1541 ele conseguiu. Quando a notícia se espalhou, foi organizada uma competição matemática entre Tartaglia e Fior. Nesta competição, cada um tinha que propor trinta

1 Scipione del Ferro foi um matemático italiano, que nasceu na Bolonha, em 6 de fevereiro de 1465 e faleceu na mesma cidade em 5 de novembro de 1526.

2 Girolamo Cardano foi um polímata italiano. Escreveu mais de 200 trabalhos sobre medicina, matemática, física, filosofia, religião e música. Nasceu em *Pavia*, no dia 24 de Setembro de 1501 e faleceu em *Roma*, no dia 21 de setembro de 1576.

3 Tartaglia é um pseudônimo do matemático italiano Niccolo Fontana, nascido em *Bréscia*, c.1500 e falecido em *Veneza*, no dia 13 de dezembro de 1557. Quando criança tinha recebido um corte de sabre, na tomada de Bréscia pelos franceses em 1512, e isso lhe prejudicou a fala. Por esse fato é que recebeu o apelido de Tartaglia, ou gago, nome que usou em lugar do de Niccolo Fontana que recebera ao nascer [2].

problemas para que o outro resolvesse em um determinado período de tempo. Quando chegou o dia da decisão, Tartaglia tinha resolvido todos os problemas propostos por Fior, enquanto que Fior não resolveu, sequer um problema proposto por Tartaglia [2]. Já em [1] é contado que Tartaglia descobre como resolver equações cúbicas da forma $x^3 + px + q$ um pouco antes do término da competição mencionada, em 13 de fevereiro de 1535.

A *Ars Magna* de Cardano, foi um progresso extraordinário que causou grande impacto entre os algebristas naquele ano e por conta disto é muitas vezes tomado como marco inicial do período moderno na matemática [2]. Contudo, a descoberta da solução quer da cúbica, quer da quártica não foi de Cardano. Ele próprio admitiu em seu livro que foi Tartaglia [4], que lhe tinha fornecido a uma sugestão para resolver a cúbica. Quanto a quártica, esta tinha sido descoberta primeiramente por um antigo amanuense[5] de Cardano, Ludovico Ferrari [6].

O que Cardano não contou foi que Tartaglia lhe contou em segredo, pois o próprio Tartaglia pretendia firmar sua reputação publicando a solução da cúbica e assim coroar o seu tratado sobre álgebra. Tartaglia, também não foi ético, publicou uma tradução de Arquimedes[7] (1543), derivada de Moerbeke[8], dando a impressão de que era obra sua, e em seu *Quesiti et inventioni diverse* (Veneza, Itália, 1546) ele deu a lei do plano inclinado, presumivelmente derivada de Jordanus Nemorarius[9], sem atribuição apro-

4 Tartaglia é um pseudônimo do matemático italiano Niccolo Fontana, nascido em *Bréscia*, c.1500 e falecido em *Veneza*, no dia 13 de dezembro de 1557. Quando criança tinha recebido um corte de sabre, na tomada de Bréscia pelos franceses em 1512, e isso lhe prejudicou a fala. Por esse fato é que recebeu o apelido de Tartaglia, ou gago, nome que usou em lugar do de Niccolo Fontana que recebera ao nascer [2].

5 Carl B. Boyer em [2] refere-se à Ludovico Ferrari neste trecho simplesmente como *amanuense* e amanuense ou copista pelo dicionário é aquele que copia textos ou documentos à mão

6 Lodovico Ferrari foi um matemático italiano, que começou a carreira como auxiliar de Cardano. Nascido em *Milão*, no dia 2 de fevereiro de 1522 e faleceu no dia 5 de outubro de 1565.

7 Arquimedes de Siracusa (287 a.C. - 212 a.C.) foi um matemático, físico, engenheiro, inventor, e astrônomo grego. Embora poucos detalhes de sua vida sejam conhecidos, são suficientes para que seja considerado um dos principais cientistas da Antiguidade Clássica.

8 Willem van Moerbeke, conhecido em espanhol como Guillermo de Moerbeke (c.1215 - c.1286) foi um tradutor medieval prolífico de textos filosóficos, médicos e científicos, famoso por seu trabalho na adaptação grega em textos latinos. Sua obra foi muito influente em sua época, dada a dificuldade de acesso às fontes originais e foi um apoio importante para o desenvolvimento da filosofia medieval, particularmente para o tomismo . Ainda hoje as suas obras são levados em consideração pelos estudiosos atuais.

9 Jordanus de Nemore (c.1225 - c.1260), também conhecido como Jordanus Nemorarius e Giordano de Nemi, foi um matemático europeu do século XIII e cientista. Ele escreveu tratados em pelo menos 6 diferentes importantes assuntos matemáticos; a ciência de pesos, “algorismi” tratados de aritmética prática, aritmética pura, álgebra, geometria e projeção estereográfica. A maioria destes tratados existem

priada. Talvez Tartaglia tenha recebido uma sugestão para a solução da cúbica de uma fonte mais antiga. Esta é uma história bem controversa. [2]

2.1 MÉTODO DE CARDANO-TARTAGLIA

Naquela época, as equações eram dividas em tipos. O método descrito por Cardano, usavam equações com coeficientes numéricos específicos como representantes de categorias gerais. Por exemplo, quando escrevia, "Seja o cubo e seis vezes o lado igual a 20" (ou $x^3 + 6x = 20$), ele evidentemente estava pensando nessa equação como tipica de todas as que têm "um cubo e coisa igual a um número", i.e., da forma $x^3 + px = q$. A solução dessa equação cobre um par de páginas de retórica. [2]

Para nos situarmos um pouco sobre a época em que estes problemas foram resolvidos, a matemática Europeia teve um longo período de estagnação em relação ao conhecimento adquirido em períodos anteriores de atividades dos árabes e dos gregos antigos. Os algarismos arábicos tinham sido introduzidos na Europa, que foi útil para os cálculos usados no comércio. E embora a matemática fosse usada em muitas aplicações comerciais, os números negativos permaneceram desconhecidos. Além disso, a notação matemática era lenta para se desenvolver. Assim por exemplo, no século XV a notação R3 V31 m R16 foi utilizada para a expressão $\sqrt[3]{31 - \sqrt{16}}$ [1].

Passaremos agora a descrever, de maneira moderna, como Cardano resolvia equação do tipo:

$$t^3 + pt + q = 0. \tag{2.1}$$

Para isto, vou reescrever a equação (2.1), tomando duas variáveis u e v, tais que:

$$u + v = t. \tag{2.2}$$

Assim, substituindo t da equação (2.2) na equação (2.1), obtemos:

em várias versões ou regravações da Idade Média. Não sabemos nada sobre ele pessoalmente, que não seja a data aproximada de sua obra.

$$\begin{aligned}(u+v)^3 + p(u+v) + q &= 0\\ u^3 + 3u^2v + 3uv^2 + v^3 + p(u+v) + q &= 0\\ u^3 + v^3 + 3uv(u+v) + p(u+v) + q &= 0\\ u^3 + v^3 + (u+v)(3uv+p) + q &= 0. \qquad (2.3)\end{aligned}$$

Agora Cardano impõe uma condição que mostra a motivação pela qual foi substituído a variável t por $u+v$. Admitindo que existam u e v, tais que:

$$3uv + p = 0.$$

Deste modo a equação (2.3) fica:

$$u^3 + v^3 + q = 0. \qquad (2.4)$$

Lembrando que $(u+v)^3 = u^3 + 3u^2v + 3uv^2 + v^3$, podemos isolar $u^3 + v^3$ na equação junto com o fato de $p = -3uv$ e assim obter a equação (2.4) na forma reduzida:

$$\begin{aligned}(u+v)^3 - 3uv(u+v) + q &= 0\\ (u+v)^3 + p(u+v) + q &= 0.\end{aligned}$$

Isolando v de $p = -3uv$ e depois substituindo na equação (2.4), obtemos:

$$\begin{aligned}u^3 + v^3 + q &= 0\\ u^3 + \left(\frac{-p}{3u}\right)^3 + q &= 0\\ u^3 - \frac{p^3}{27u^3} + q &= 0.\end{aligned}$$

Multiplicando a equação por u^3 e fazendo $u^3 = z$, temos:

$$z^2 + qz - \frac{p^3}{27} = 0.$$

Resolvendo esta equação quadrática, vem

$$
\begin{aligned}
z &= \frac{-q \pm \sqrt{q^2 - 4 \cdot 1 \cdot \frac{-p^3}{27}}}{2 \cdot 1} \\
z &= -\frac{q}{2} \pm \frac{\sqrt{q^2 + 4 \cdot \frac{p^3}{27}}}{\sqrt{4}} \\
z &= -\frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{4p^3}{27 \cdot 4}} \\
z &= -\frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}.
\end{aligned}
$$

Como $u^3 = z$, temos $u = \sqrt[3]{z}$. Assim:

$$
u = \sqrt[3]{-\frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}.
$$

Como é possível resolver analogamente para v vou considerar uma solução u e a outra v, logo:

$$
\begin{aligned}
u &= \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}, \\
v &= \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}.
\end{aligned}
$$

Desta forma podemos determinar t e assim obter a *fórmula de Cardano* para resolver equações cúbicas do tipo $t^3 + pt + q = 0$. Lembrando que $t = u + v$, chegamos a:

$$
t = u + v = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}. \tag{2.5}
$$

Cardano também resolveu equações cúbicas que envolviam termos quadráticos, em sua *Ars Magna*. Iremos descrever, como fizemos anteriormente, como resolver uma equação do tipo:

$$
x^3 + ax^2 + bx + c = 0. \tag{2.6}
$$

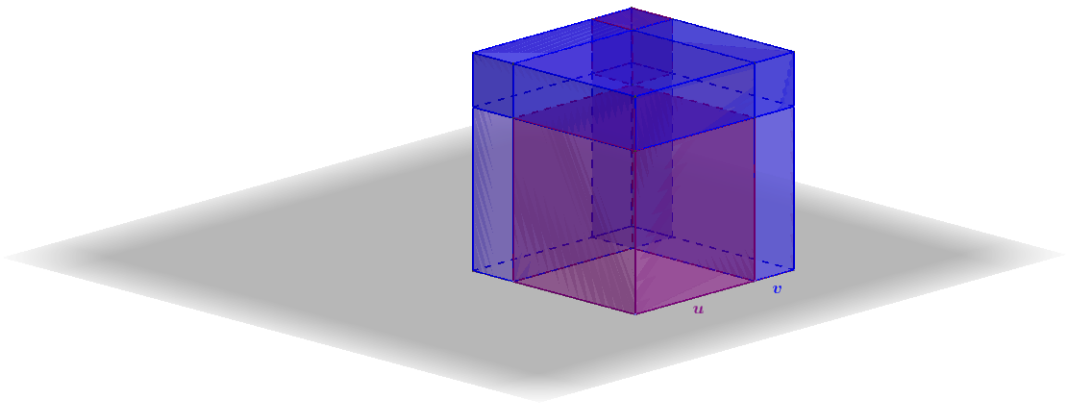

Figura 10: O cubo aqui retratado é a base geométrica da equação binomial, semelhante à apresentação de Cardano em sua *Ars Magna*. O cubo maior pode ser decomposto em dois cubos menores e 3 paralelepípedos retangulares, todos com lado de comprimentos u, v e $u+v$.

Para isto, Cardano estabelece uma maneira de relacionar as equações (2.6) e (2.1). Neste sentido, ele escreve $x^3 + ax^2$ de outro modo. Sabendo que

$$\left(x + \frac{a}{3}\right)^3 = x^3 + ax^2 + \frac{a^2}{3}x + \frac{a^3}{27},$$

podemos escrever $x^3 + ax^2$, como

$$x^3 + ax^2 = \left(x + \frac{a}{3}\right)^3 - \frac{a^2}{3}x - \frac{a^3}{27} = \left(x + \frac{a}{3}\right)^3 - \frac{a^2}{3}\left(x + \frac{a}{3}\right) + \frac{2}{27}a^3,$$

e para concluir esta etapa, substituímos todas as ocorrências de x na equação por $x = t - \frac{a}{3}$, obtendo:

$$\left(x + \frac{a}{3}\right)^3 - \frac{a^2}{3}\left(x + \frac{a}{3}\right) + \frac{2}{27}a^3 = \left(t - \frac{a}{3} + \frac{a}{3}\right)^3 - \frac{a^2}{3}\left(t - \frac{a}{3} + \frac{a}{3}\right) + \frac{2}{27}a^3 = t^3 - \frac{a^2}{3}t + \frac{2}{27}a^3.$$

Agora escrevendo as equações (2.6) e 2.1, temos

$$x^3 + ax^2 + bx + c = t^3 + pt + q,$$

portanto é imediato que

$$p = -\frac{1}{3}a^2 + b \qquad \text{e} \qquad q = \frac{2}{27}a^3 - \frac{1}{3}ab + c,$$

o que conclui o método.

Podemos perceber que não é simples resolver equações usando este método. Façamos um exemplo, para ilustrar como o método funciona.

Exemplo 2.1. Calculemos a equação $x^3 - 3x^2 - 3x - 1 = 0$.

Solução

Temos os coeficientes $a = -3$, $b = -3$ e $c = -1$. Logo

$$p = -\frac{1}{3}(-3)^2 + (-3) = -6 \quad \text{e} \quad q = \frac{2}{27}(-3)^3 - \frac{1}{3}(-3)(-3) + (-1) = -2 - 3 - 1 = -6,$$

portanto, chegamos a equação $t^3 - 6t - 6 = 0$. Como vimos, esta equação pode ser resolvida com a fórmula (2.5). Assim:

$$\begin{aligned} t &= \sqrt[3]{-\frac{(-6)}{2} + \sqrt{\frac{(-6)^2}{4} + \frac{(-6)^3}{27}}} + \sqrt[3]{-\frac{(-6)}{2} - \sqrt{\frac{(-6)^2}{4} + \frac{(-6)^3}{27}}} \\ &= \sqrt[3]{3 + \sqrt{1}} + \sqrt[3]{3 - \sqrt{1}} \\ &= \sqrt[3]{4} + \sqrt[3]{2}. \end{aligned}$$

Lembrando que $x = t - \frac{a}{3}$ e que $a = -3$, é imediato que $x = t + 1$ e assim chegamos a solução:

$$x = \sqrt[3]{4} + \sqrt[3]{2} + 1.$$

2.2 A EQUAÇÃO CÚBICA E O SURGIMENTO DOS NÚMEROS COMPLEXOS

Descrevemos com a utilização de números negativos numa linguagem moderna, como Cardano resolvia as equações cúbicas da forma $x^3 + ax^2 + bx + c = 0$. Se considerarmos uma equação cúbica $Ax^3 + Bx^2 + Cx + D = 0$ qualquer com $A \neq 0$, podemos dividir todos os termos por A e obter $x^3 + (B/A)x^2 + (C/A)x + D/A = 0$ e assim o método de Cardano parece resolver qualquer equação cúbica, bastando tomar $a = B/A$, $b = C/A$ e $c = D/A$. No entanto, Cardano se deparou com casos em que a fórmula parecia falhar, pois para alcançar o resultado era necessário considerar a existência de raízes quadradas de números negativos. Cardano utilizou esses números a princípio como valores intermediários, e mais tarde como objetos matemáticos em si mesmos, para que lá se desenvolvesse um interesse independente [1]. Os casos em que eram necessário admitir tais números foi chamado de *casus irreducibilis*. Estes casos ocorrem quando a equação cúbica da forma reduzida $x^3 + px + q = 0$, tiver o radicando da raiz quadrada negativo:

$$\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3 < 0.$$

Por exemplo, a equação $x^3 - 8x - 3 = 0$, tem claramente como solução 3, mas a fórmula nos leva a expressão complicada

$$x = \sqrt[3]{\frac{3}{2} + \frac{19}{6}\sqrt{-\frac{5}{3}}} + \sqrt[3]{\frac{3}{2} - \frac{19}{6}\sqrt{-\frac{5}{3}}}.$$

O primeiro avanço neste sentido foi dado por Rafael Bombelli [10]. No ano de sua morte foi publicado seu livro, *L'Algebra*. Neste livro ele resolve a equação $x^3 = 15x + 4$, calculando uma expressão com radical negativo. Primeiro usando a fórmula de Cardano obtemos a expressão:

$$x = \sqrt[3]{-\frac{-4}{2} + \sqrt{\frac{(-4)^2}{4} + \frac{(-15)^3}{27}}} + \sqrt[3]{-\frac{-4}{2} - \sqrt{\frac{(-4)^2}{4} + \frac{(-15)^3}{27}}} = \sqrt[3]{2 + 11\sqrt{-1}} + \sqrt[3]{2 - 11\sqrt{-1}}.$$

Nos casos em que aparecia raízes quadradas de números negativos, Cardano se referia a esses números como "*sofisticas*" e concluía que o resultado nesse caso era "tão sutil quanto inútil". Cardano tem o mérito de ter dado atenção a este intrigante problema [2][11].

Voltando à Bombelli, ele chamou "ideia louca", pois toda a questão "parecia apoiar-se em sofismas" [2], ele comentou "Um pensamento extravagante, de acordo com muitos, me foi por um longo tempo de mesma opinião. O assunto parecia descansar mais em sofismas que na verdade, mas eu procurei até que encontrei uma prova" [1][12].

Ele efetuou corajosamente para equação $x^3 = 15x + 4$, cálculos com expressões em que haviam radicais com números negativos. Ele obteve a raiz cúbica de $\sqrt[3]{2 + 11\sqrt{-1}}$, fazendo o cálculo de $\left(2 + \sqrt{-1}\right)^3$, para isto, fez

$$\left(2 + \sqrt{-1}\right)^3 = 8 + 12\sqrt{-1} - 6 - \sqrt{-1} = 2 + 11\sqrt{-1},$$

e de modo semelhante calculou a raiz cúbica de $\sqrt[3]{2 - 11\sqrt{-1}}$, observando que

$$\left(2 - \sqrt{-1}\right)^3 = 8 - 12\sqrt{-1} - 6 + \sqrt{-1} = 2 - 11\sqrt{-1},$$

10 Rafael Bombelli foi um matemático e engenheiro hidráulico italiano. Nasceu na Bolonha em 1526 e faleceu em Roma no ano de 1572.

11 Nota de rodapé de [2]"Não foi publicada nenhuma tradução de toda a *Ars magna* mas uma seleção aparece em D. E. Smith, *A Source Book in Mathematics* (1929). Numa comunicação recente D. J. Struik informou que existe em manuscrito uma tradução para o inglês da *Ars magna* por J. R. Witner em Washington. Deve ser publicada pela M. L. T. Press."

12 Nota de rodapé de [1] Citado por Moritz Cantor, *Vorlesungen über Geschichte der Mathematik*, Berlim, 1900-1908, Band 2, p. 625.

assim, somando as duas raízes chegamos ao conhecido resultado 4, pois $x = 2 + \sqrt{-1} + 2 - \sqrt{-1} = 4$.

Esses cálculos ousados tinham fornecido uma explicação para um resultado já sabido. É certo, que na história da matemática, evolução semelhante ocorreu quando números negativos primeiro eram vistos como um complemento útil para o conjunto de números admissíveis. Entretanto ao compararmos com os números negativos, a raiz quadrada de números negativos possui um nível de abstração bem maior, uma vez que não há uma analogia em nossa vida cotidiana que seja próxima. Ao contrário dos números negativos que comparamos com um saldo negativo de uma conta bancária, ou mesmo a simples ideia de ficar devendo dinheiro, por exemplo.

Assim levou quase mais duzentos anos antes de que os objetos hesitantes introduzidos por Bombelli serem aceitos em uso na matemática geral, sob o nome de números complexos. O que era necessário era uma descrição de suas propriedades fundamentais para que não houvesse dúvida de como eles poderiam ser usados. Isso ocorreu quando a questão "o que realmente são números complexos?" foi posto de lado em favor de defini-los a partir de suas propriedades, o que foi feito primeiro em 1797 Caspar Wessel[13]. No entanto, a definição formal de Wessel de nenhuma maneira eliminou todas as dúvidas, até porque seus escritos não foram largamente divulgados. Assim foi por quase um meio século que esses números imaginários ou impossíveis não foram amplamente aceitos [1].

2.2.1 *Números Complexos*

Representação Cartesiana

Podemos definir o conjunto dos *números complexos* $\mathbb{C}$, como todos os pares (a, b) cujas coordenadas a e b são números reais. Geometricamente, o conjunto de números complexos pode ser visto como um plano, em analogia com a reta que representa o conjunto dos números reais. Representamos o número complexo associado ao par ordenado (a, b), por

$$a + b\sqrt{-1}.$$

13 Caspar Wessel foi um matemático dinamarquês-norueguês que descobriu em 1797 uma representação gráfica para os números complexos, publicada em 1798 nas atas da academia dinamarquesa. Nascem em *Vestby*, no dia 8 de junho de 1745 e faleceu em Copenhague, no dia 25 de março de 1818.

Definimos as operações (+ soma e $\times$ multiplicação) matemáticas dos números complexos, onde $a, b, c, d \in \mathbb{R}$ como:

$$(a, b) + (c, d) = (a + c, b + d),$$
$$(a, b) \times (c, d) = (ac - bd, bc + ad).$$

Deste modo,

$$a + b\sqrt{-1} + c + d\sqrt{-1} = (a + c) + (b + d)\sqrt{-1},$$
$$\left(a + b\sqrt{-1}\right) \times \left(c + d\sqrt{-1}\right) = (ac - bd) + (bc + ad)\sqrt{-1}.$$

As operações inversas são definidas utilizando os elementos inversos (ou simétricos). Assim, a subtração é definida como a adição de um valor negativo e a divisão como a multiplicação por seu inverso. Lembrando que o inverso aditivo de um número a é o número que somado a a resulta no elemento neutro 0. E o elemento neutro é o número que somado a qualquer outro, não altera o resultado, ou seja, $a + 0 = a$.

Assim o elemento neutro de $\mathbb{C}$ é $(0, 0)$, pois

$$(a, b) + (c, d) = (a, b) \iff (c, d) = (0, 0),$$

desta forma, definimos o inverso aditivo (ou oposto) de (a, b) como sendo

$$-(a, b) = (-a, -b).$$

De maneira semelhante, definimos a divisão. O elemento identidade (ou neutro) para multiplicação, de um número complexo é $(1, 0)$, pois

$$(a, b) \times (1, 0) = (1a - 0b, 1b + 0a) = (a, b),$$

assim, o inverso de (a, b) é o número que multiplicado por (a, b) resulta $(1, 0)$. Assim denotamos o inverso de (a, b) por $(a, b)^{-1}$, onde

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

com $(a, b) \neq (0, 0)$, pois

$$(a, b) \times (a, b)^{-1} = (a, b) \times \left(\frac{a}{a^2 + b^2}, \frac{-b}{a^2 + b^2}\right) = \left(a\frac{a}{a^2 + b^2} - b\frac{-b}{a^2 + b^2}, b\frac{a}{a^2 + b^2} + a\frac{-b}{a^2 + b^2}\right) = (1, 0).$$

Essas definições cumprem a meta desejada, uma vez que, exceto por não ter uma relação de ordem, os números complexos têm todas as leis conhecidas da operação dos números reais.

Podemos fazer algumas observações:

- O subconjunto dos números complexos da forma $(a, 0)$ pode ser identificado com o conjunto dos números reais nas operações, assim como o conjunto de frações com denominador 1 pode ser identificado com os números inteiros. Os números complexos, portanto, pode ser vistos como uma extensão de números reais. Por simplicidade podemos escrever o número complexo da forma $(a, 0)$ simplesmente como a. No número complexo (a, b), chamamos a de *parte real*.
- Temos $(0, 1) \times (0, 1) = (0, -1) \times (0, -1) = (-1, 0)$, um resultado que corresponde ao número real -1. Por conseguinte, os dois números complexos $(0, 1)$ e $(0, -1)$ podem ser interpretados como raízes quadradas de -1. Ao número $(0, 1)$ é dada a notação especial i, chamada de *unidade imaginária*. Em um número complexo (a, b), chamamos b de *parte imaginária*.
- Temos a equação $(a, b) \times (a, -b) = a^2 + b^2$ onde $(a, -b)$ chama-se o *conjugado* do número complexo (a, b). Isso é indicado por $\overline{(a, b)}$. Chamamos de $\sqrt{a^2 + b^2}$ o *valor absoluto* ou *módulo* do número (a, b). Situado no *plano complexo*, a representação geométrica dos números complexos é o módulo de um número complexo e representa a distância do número à origem[14]. Um exemplo é exibido na Figura 11. Finalmente, a conjugação complexa (tomar o conjugado de um número complexo) possui a seguinte propriedade: $\overline{((a, b) \times (c, d))} = \overline{(a, b)} \times \overline{(c, d)}$.

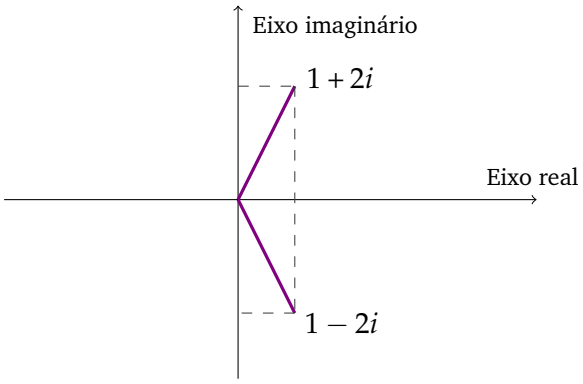

Figura 11: O plano complexo, com o número $1 + 2i$ e seu conjugado $1 - 2i$. O módulo de ambos os números é $\sqrt{5}$.

Todas estas propriedades em conjunto fazem-nos a certeza de que com pares (a, b), da forma $(a, b) = (a, 0) + (b, 0) \times (0, 1) = a + bi$, onde $i^2 = -1$, temos de fato definido

14 Nota de rodapé de [1] A definição da distância entre dois números complexos como o módulo de sua diferença torna possível a criação de uma teoria da função, ou análise complexa, sobre os números complexos, em que noções como a convergência, continuidade, derivada e integral, são definidas com propriedades semelhantes aqueles de análise clássica.

o conjunto de objetos matemáticos da forma $a + b\sqrt{-1}$. Esta extensão resultante dos números reais foi conseguida sem usar a expressão anteriormente indefinida $\sqrt{-1}$, cuja utilização, por outro lado, nem sempre é completamente sem problemas, uma vez que pode facilmente levar a cálculos errôneos, tais como $\sqrt{-1}\sqrt{-1} = \sqrt{(-1)(-1)} = \sqrt{1} = 1$ [1].

Representação Polar

Outra maneira de representar um número complexo é geometricamente. Para isto, observamos que cada número complexo localiza-se no círculo unitário, ou seja, o raio de círculo com um centro na origem, pode ser representado em termos das funções trigonométricas seno e cosseno. Para ser mais preciso, tais números complexos têm uma representação

$$\cos\phi + i\operatorname{sen}\phi,$$

em que ϕ é o ângulo que gira para a esquerda do eixo horizontal positivo (isto é, o eixo real positivo) a linha a partir da origem para o número complexo em questão (ver Figura 12).

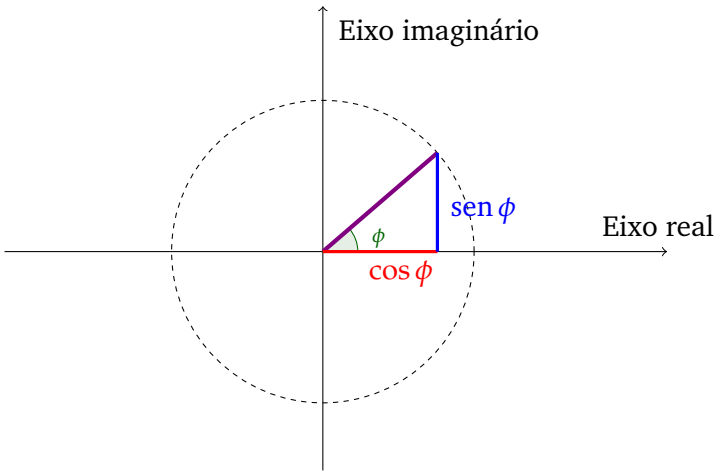

Figura 12: Representação de um número complexo da forma $\cos\phi + i\operatorname{sen}\phi$ localizado no círculo unitário.

Agora se multiplicam em conjunto dois desses números encontramos o resultado no círculo unitário, podemos fazer isto simplesmente adicionando seus ângulos. A prova disto segue facilmente das leis de adição para seno e cosseno:

$$\begin{aligned}(\cos\phi + i\operatorname{sen}\phi)(\cos\psi + i\operatorname{sen}\psi) &= (\cos\phi\cos\psi - \operatorname{sen}\phi\operatorname{sen}\psi) + i(\cos\phi\operatorname{sen}\psi + \operatorname{sen}\phi\cos\psi)\\ &= \cos(\phi+\psi) + i\operatorname{sen}(\phi+\psi).\end{aligned}$$

Podemos generalizar este resultado para todos os números complexos diferentes de zero. Introduzindo o módulo de um número complexo, que é 1 para os números complexos que se encontram no círculo unitário. Note que, se um número complexo diferente de zero de módulo m, podemos definir $s = \ln m$ (o logaritmo natural) e escrever $m = e^s$. Assim podemos definir um número complexo z com ângulo ϕ e módulo e^s, e escrever

$$z = e^s(\cos\phi + i\text{sen}\,\phi).$$

Vemos [15], então, que

$$e^s(\cos\phi + i\text{sen}\,\phi) \times e^t(\cos\psi + i\text{sen}\,\psi) = e^{s+t}\left(\cos(\phi+\psi) + i\text{sen}\,(\phi+\psi)\right).$$

O caso especial de elevar um número complexo a uma potência é determinada a partir da *fórmula de Moivre*, que apesar de levar seu nome Abraham de Moivre[16] não formulou explicitamente [1]

$$\left(e^s(\cos\phi + i\text{sen}\,\phi)\right)^n = \left(e^s\right)^n\left(\cos(n\phi) + i\text{sen}\,(n\phi)\right).$$

2.3 RESOLUÇÃO DE EQUAÇÕES UTILIZANDO NÚMEROS COMPLEXOS

Vamos iniciar este estudo, utilizando o conhecimento adquirido para resolver a equação $x^3 - 1 = 0$. Claramente $x_1 = 1$ é a única solução real. Entretanto, ao considerarmos os números complexos, a fórmula de Moivre sugere que a equação deve ter duas soluções adicionais. Se considerarmos o número complexo da forma (a, b) onde a é uma coordenada da abscissa (eixo real) e b é uma coordenada da ordenada (eixo imaginário), ambas soluções pertencem ao círculo unitário, ou seja, enxergando (a, b) como um ponto P, P é um ponto da circunferência. Na Figura 13, observamos as soluções da equação $x^3 - 1 = 0$.

15 Nota de rodapé de [1] A razão para a validade desta equação ficará claro quando a série de potência para seno, cosseno e funções exponenciais são estendidos para os números complexos, que foi realizado pela primeira vez em 1748 por Leonhard Euler (1707-1783). Então pode-se ver que para números complexos arbitrários $x + iy$, temos a identidade $e^{x+iy} = e^x(\cos y + i\text{sen}\,y)$.

16 Abraham de Moivre foi um matemático francês famoso pela Fórmula de De Moivre, que relaciona os números complexos com a trigonometria, e por seus trabalhos na distribuição normal e na teoria das

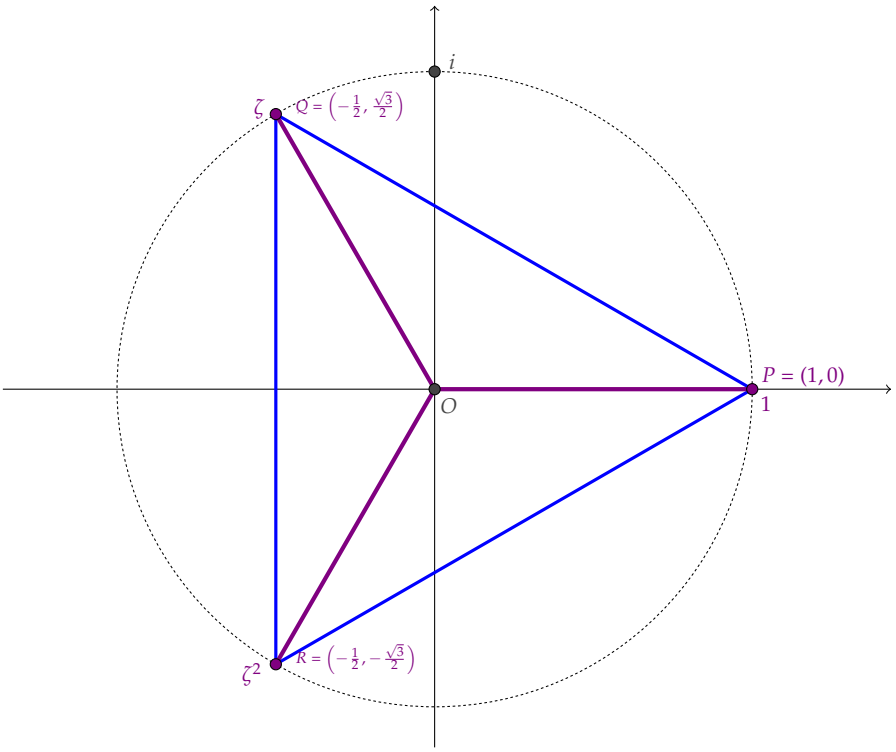

Figura 13: As três soluções $1, \zeta$ e ζ^2 da equação $x^3 - 1 = 0$, que correspondem aos pontos P, Q e R.

Podemos escrever a solução x_2 como sendo $\zeta = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$ na forma $\zeta = \cos\left(\frac{2\pi}{3}\right) + i\text{sen}\left(\frac{2\pi}{3}\right)$, onde $m = e^s = 1$. Utilizando a fórmula de Moivre, obtemos

$$\zeta^3 = (1)^3\left(\cos\left(3\frac{2\pi}{3}\right) + i\text{sen}\left(3\frac{2\pi}{3}\right)\right) = \cos(2\pi) + i\text{sen}\,(2\pi) = 1.$$

Assim, $\zeta^3 = 1$.

Analogamente, podemos escrever a solução x_3, também em função de ζ, pois $\zeta^2 = -\frac{1}{2} - \frac{\sqrt{3}}{2}i$ uma vez que

$$\zeta^2 = \cos\left(2\frac{2\pi}{3}\right) + i\text{sen}\left(2\frac{2\pi}{3}\right) = \cos\left(\frac{4\pi}{3}\right) + i\text{sen}\left(\frac{4\pi}{3}\right),$$

desta forma

$$(\zeta^2)^3 = (1)^3\left(\cos\left(3\frac{4\pi}{3}\right) + i\text{sen}\left(3\frac{4\pi}{3}\right)\right) = \cos(4\pi) + i\text{sen}\,(4\pi) = 1,$$

probabilidades. Nasceu em Champagne, na França, no dia 26 de Maio de 1667 e faleceu em Londres, no Reino Unido, no dia 27 de Novembro de 1754.

logo $(\zeta^2)^3 = 1$.

Assim as três soluções para a equação $x^3 - 1 = 0$ são

$$x_1 = 1, \qquad x_2 = -\frac{1}{2} + \frac{\sqrt{3}}{2}i \qquad \text{e} \qquad x_3 = -\frac{1}{2} - \frac{\sqrt{3}}{2}i.$$

2.3.1 *A Extensão da Fórmula de Cardano*

Equações do tipo $x^n - 1 = 0$ são chamadas de *equações ciclotômicas*, o leitor pode encontrar mais em [1]. Com as raízes dessas equações quando $n = 3$ podemos estender a fórmula de Cardano para que as três raízes sejam sempre obtidas. Lembremos que usamos duas equações auxiliares para resolver pela fórmula de Cardano,

$$3uv = -p \qquad \text{e} \qquad u^3 + v^3 = -q.$$

Para $x^3 - 1 = 0$, temos $p = 0$ e $q = -1$, utilizando a fórmula geral, temos:

$$x = \underbrace{\sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}}_{u} + \underbrace{\sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}}_{v} = \sqrt[3]{-\frac{-1}{2} + \sqrt{\frac{(-1)^2}{4}}} + \sqrt[3]{-\frac{-1}{2} - \sqrt{\frac{(-1)^2}{4}}} = 1 + 0 = 1.$$

Se considerarmos a solução $x_2 = \zeta = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$, podemos obter uma expressão para a solução de x_2 simplesmente multiplicando v por ζ, uma vez que em nosso problema $u = 0$. E de modo semelhante podemos multiplicar v por ζ^2, para obter a solução $x_3 = \zeta^2 = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$. Mas e quanto a u, deve ser multiplicado por algum número? Para responder esta questão, podemos observar que no *casus irreducibilis*, os dois números complexos u^3 e v^3 formam um par de números conjugados, o mesmo ocorre com ζ e ζ^2. O que nos sugere por quais números nas expressões para x_2 e x_3 devemos multiplicar v. Assim, as três expressões para resolver uma equação qualquer da forma $x^3 + px + q = 0$, são:

$$x_1 = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}},$$
$$x_2 = \zeta\sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} + \zeta^2\sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}},$$
$$x_3 = \zeta^2\sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} + \zeta\sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}.$$

Destacamos que este resultado nos aponta uma direção curiosa. Se uma equação tiver três soluções reais, a fórmula de Cardano sem a utilização dos números complexos, não nos levará a nenhuma das soluções. Uma vez que, no caso geral de três soluções x_1, x_2 e x_3 reais, teremos

$$\overline{x_j} = \overline{\zeta^{j-1}u + \zeta^{-(j-1)}v} = \zeta^{-(j-1)}\overline{u} + \zeta^{j-1}\overline{v} = \zeta^{-(j-1)}v + \zeta^{j-1}u = x_j,$$

para $j = 1, 2, 3$.

No entanto estas expressões são de difícil resolução, note este exemplo retirado do livro [1]. O problema $x^3 = 8x + 3$, que aparece no início da *Ars Magna* de Cardano, que obtém a solução de

$$\begin{aligned} x_1 &= \sqrt[3]{\frac{3}{2} + i\frac{19}{6}\sqrt{\frac{5}{3}}} + \sqrt[3]{\frac{3}{2} - i\frac{19}{6}\sqrt{\frac{5}{3}}} \\ &= \frac{1}{2}\left(3 + i\sqrt{\frac{5}{3}}\right) + \frac{1}{2}\left(3 - i\sqrt{\frac{5}{3}}\right) \\ &= 3. \end{aligned}$$

As outras duas soluções, que Cardano sabia, são

$$\begin{aligned} x_2 &= \frac{1}{4}\left(-1 + i\sqrt{3}\right)\left(3 + i\sqrt{\frac{5}{3}}\right) + \frac{1}{4}\left(-1 - i\sqrt{3}\right)\left(3 - i\sqrt{\frac{5}{3}}\right) \\ &= \frac{1}{2}\left(-3 - \sqrt{5}\right) \end{aligned}$$

e

$$
\begin{aligned}
x_3 &= \frac{1}{4}\left(-1 - i\sqrt{3}\right)\left(3 + i\sqrt{\frac{5}{3}}\right) + \frac{1}{4}\left(-1 + i\sqrt{3}\right)\left(3 - i\sqrt{\frac{5}{3}}\right) \\
&= \frac{1}{2}\left(-3 + \sqrt{5}\right).
\end{aligned}
$$

Assim, apesar do método conduzir a solução, não é nada simples utilizar estas fórmulas. Existem métodos numéricos que conduzem a resultados aproximados com mais rapidez e com boa precisão. No Capítulo 7, apresentamos um destes métodos.

3

EQUAÇÕES QUÁRTICAS

Como dissemos no capítulo anterior, foi Ludovico Ferrari que mostrou a Cardano uma maneira de resolver uma equação quártica. Cardano escreveu na *Ars magna* que "é devida a Ludovico Ferrari, que a inventou a meu pedido". Ferrari, em notação atual, resolveu por exemplo a equação $x^4 + 6x^2 + 36 = 60x$. A solução descrita por Cardano, foi retirada de [2]. Ele procedeu da seguinte maneira:

1. Primeiro somar suficientes quadrados e números a ambos os lados para que o primeiro membro fique um quadrado perfeito, nesse caso $x^4 + 12x^2 + 36$ ou $(x^2 + 6)^2$. Assim somamos, $6x^2$ a ambos os lados.

$$x^4 + 6x^2 + 36 = 60x$$
$$x^4 + 12x^2 + 36 = 6x^2 + 60x.$$

2. Agora somar a ambos os membros da equação termos envolvendo uma nova incógnita y de modo que o primeiro membro permaneça um quadrado perfeito, como $(x^2 + 6 + y)^2$. Mas $(x^2 + 6 + y)^2 = y^2 + 2x^2y + 12y + x^4 + 12x^2 + 36$ e como $x^4 + 12x^2 + 36 = 6x^2 + 60x$. A equação agora fica

$$\begin{aligned}(x^2 + 6 + y)^2 &= 6x^2 + 60x + y^2 + 12y + 2yx^2,\\ &= (2y + 6)x^2 + 60x + (y^2 + 12y).\end{aligned}$$

3. O passo crucial seguinte consiste em escolher y do modo que o trinômio no segundo membro fique um quadrado perfeito. Isso se faz, é claro, igualando a zero o discriminante Δ, uma regra antiga e bem conhecida. Sabendo que $\Delta = b^2 - 4ac$, obtemos:

$$\begin{aligned} 60^2 - 4(2y+6)(y^2+12y) &= 0 \\ 3600 - 8y^3 - 120y^2 - 144y &= 0 \\ \frac{3600 - 8y^3 - 120y^2 - 144y}{8} &= \frac{0}{8} \\ 450 - y^3 - 15y^2 - 36y &= 0. \end{aligned}$$

4. Do passo 3 resulta uma equação cúbica em y, $y^3 + 15y^2 + 36y = 450$, hoje chamada a "cúbica resolvente"[1] da equação quártica dada. Essa é agora resolvida. em relação a y pelas regras previamente dadas[2] para resolução de equações cúbicas, sendo o resultado

$$y = \sqrt[3]{\frac{287}{2} + \sqrt{\frac{80449}{4}}} + \sqrt[3]{\frac{287}{2} - \sqrt{\frac{80449}{4}}} - 5.$$

5. Substituir o valor de y obtido em 4 na equação para x do passo 2 e extrair a raiz quadrada de ambos os membros.

6. O resultado do passo 5 é uma equação quadrática, que deve agora ser resolvida a fim de achar o valor de x desejado.

3.1 MÉTODO DE RESOLUÇÃO DE EQUAÇÕES QUÁRTICAS

As equações quárticas que foram descritas por Cardano em sua *Ars Magna*, graças a seu aluno, Ludovico Ferrari, são chamadas de biquadradas. Estas equações da forma $x^4 + px^2 + qx + r = 0$ foram resolvidas usando um procedimento similar ao usado para resolver a equação $x^3 + px + q = 0$, que é transformá-la em uma equação do tipo que saibamos resolver. E em linhas gerais, foi isto que foi feito.

Tomando a equação $x^4 + px^2 + qx + r = 0$ adicionamos dois termos de potências de x e x^2, de tal forma que um quadrado perfeito seja obtido em ambos os lados da equação. Em seguida adicionamos $2zx^2 + z^2$ para ambos os lados da equação, para assim obter

$$x^4 + 2zx^2 + z^2 = (2z - p)x^2 - qx + (z^2 - r).$$

1 Boyer chama de cúbica resolvente a equação da forma $x^3 + px + q = 0$. Outros autores a chamam de equação cúbica reduzida.

2 *As regras previamente citadas* do livro [2], se trata dos procedimentos que levam à fórmula (2.5), pelo qual a equação pode ser resolvida.

Embora o lado esquerdo da equação já esteja na forma de um quadrado perfeito, $(x^2+z)^2$, o mesmo não ocorre necessariamente para o lado direito, no entanto, z pode agora ser convenientemente escolhido, de tal modo, que satisfaça a condição

$$2\sqrt{2z-p}\sqrt{z^2-r} = -q.$$

A quadratura de ambos os lados dessa condição leva a

$$(2z-p)\left(z^2-r\right) = \frac{q^2}{4},$$

que por sua vez, nos leva a equação cúbica

$$z^3 - \frac{p}{2}z^2 - rz + \frac{pr}{2} - \frac{q^2}{8} = 0.$$

A escolha de z não foi aleatória, pois sabemos resolver a cúbica usando a equação (2.5). Com isso, podemos agora escrever

$$x^2 + z = \pm\left(\sqrt{2z-p}x + \sqrt{z^2-r}\right),$$

onde cada um dos dois possíveis sinais produz duas soluções em virtude da fórmula quadrática. No total, portanto, obtém as seguintes quatro soluções:

$$x_{1,2} = \frac{1}{2}\sqrt{2z-p} \pm \sqrt{-\frac{1}{2}z - \frac{1}{4}p + \sqrt{z^2-r}},$$
$$x_{3,4} = \frac{1}{2}\sqrt{2z-p} \pm \sqrt{-\frac{1}{2}z - \frac{1}{4}p - \sqrt{z^2-r}}.$$

De modo semelhante ao que foi feito às equações cúbicas completas, foi feito às quárticas completas. Ou seja, a resolução da equação quártica completa na forma

$$x^4 + ax^3 + bx^2 + cx + d = 0,$$

é feita transformando esta equação em uma do tipo

$$y^4 + py^2 + qy + r = 0.$$

Em analogia ao caso da equação cúbica, isto pode ser feito substituindo a variável x através da substituição

$$x = y - \frac{a}{4},$$

que elimina o termo cúbico e assim chegando na forma desejada:

$$x^4 + ax^3 + bx^2 + cx + d = y^4 + py^2 + qy + r.$$

Evidentemente, como fizemos no caso das equações cúbicas, os coeficientes da equação reduzida podem ser calculados da equação original usando expressões polinomiais.

4

GRUPOS, ANÉIS E CORPOS

Neste capítulo vamos desenvolver uma série de conceitos algébricos muito importantes para a Matemática e sobretudo para a Álgebra. Forneceremos aqui a base para a futura resolução do problema de encontrar para um dado polinômio $p(x)$, com coeficientes num corpo F, mas sem raízes em F, um corpo K, extensão de F, onde $p(x)$ tenha raiz. A construção de tal K envolve as seguintes estruturas:

- F: um corpo. Um exemplo de corpo é o conjunto dos inteiros módulo um número primo, que é indicado por $\mathbb{Z}_p$.
- $F[x]$: o conjunto dos polinômios na variável x e coeficientes no corpo F.
- $p(x)$: um polinômio de $F[x]$.
- $(p(x))$: o ideal de $p(x)$ em $F[x]$.
- $K = F[x]/(p(x))$: um corpo dado por um quociente cujos elementos são classes de polinômios com mesmo resto na divisão por $p(x)$.

4.1 CONCEITOS PRELIMINARES

Vamos inicialmente, introduzir alguns conceitos que serão de grande valia para o estudo que estamos iniciando neste capítulo. Começaremos definindo o conceito de *relação de equivalência*.

Definição 4.1. A relação binária, $\sim$, sobre A é dita uma *relação de equivalência* sobre um conjunto A se para quaisquer a, b e c em A:

1. $a \sim a$ (*reflexividade*);
2. $a \sim b$ implica $b \sim a$ (*simetria*);
3. $a \sim b$ e $b \sim c$ então $a \sim c$ (*transitividade*).

É fácil perceber que a igualdade é uma relação de equivalência. Pois:

- $a = a$ (*reflexividade*);
- $a = b$ implica $b = a$ (*simetria*);
- $a = b$ e $b = c$ então $a = c$ (*transitividade*).

Vejamos outro exemplo de relação de equivalência.

Exemplo 4.1. Sejam $a, b \in \mathbb{Z}$ definimos $a \sim b$ se $a - b$ for par. Façamos a verificação:

1. $a - a = 0$ e 0 é par. Assim $a \sim a$.
2. $a \sim b$ implica que $a - b$ é par, assim $b - a$ é par, pois $b - a = -(-b + a) = -(a - b)$ que evidentemente é par. Portanto, $b \sim a$.
3. $a \sim b$ e $b \sim c$ então $a - b$ e $b - c$ são inteiros pares, o que implica que $a - c = (a - b) + (b - c)$ também é par e portanto $a \sim c$.

Outra relação de equivalência muito importante é a *congruência modular* que definimos:

Definição 4.2. Se a e b são inteiros e m é um inteiro positivo, então a é *congruente com b módulo m*, indicado por $a \equiv b \mod m$ se m divide $a - b$ que indicamos por $m \mid a - b$.

Notemos que o Exemplo 4.1 se trata de todos os elementos que são congruentes a 0 módulo 2. Vamos definir um conjunto, onde reunamos todos os elementos em $\mathbb{Z}$ que sejam congruentes a um certo inteiro a módulo n. Este conjunto é denominado de *classe de equivalência*, que neste caso, também pode ser chamado de *classe de congruência* .

Definição 4.3. Considere a um número inteiro. Indicamos por $[a]$, a *classe de equivalência*, ou neste caso, a *classe de congruência* de a módulo n, o conjunto formado pelos números que deixam o mesmo resto na divisão por n.

Da maneira como definimos classes de congruência, notamos que existem n classes distintas. Isto pode ser verificado, pois pelo algoritmo de Euclides, temos $a = kn + r$, onde $0 \leq r < n$, logo $a \equiv r \mod n$. E consequentemente, as n classes de congruência distintas são: $[0]$, $[1]$, ..., $[n-1]$. Fato importante, é que estas classes de equivalências são disjuntas.

Teorema 4.4. *As classes de equivalência distintas de uma relação de equivalência sobre um conjunto A nos fornecem uma decomposição de A como uma reunião de subconjuntos mutuamente disjuntos. Reciprocamente, dada uma decomposição de um conjunto A como uma reunião de subconjuntos mutuamente disjuntos e não vazios, podemos definir uma relação de equivalência sobre A para a qual estes subconjuntos sejam as classes de equivalência distintas.*

O Teorema 4.4 está demonstrado em [6].

Agora que já definimos um ponto de partida voltemos um pouco aos elementos listados. O conjunto dos inteiros módulo um número n que denotamos por $\mathbb{Z}_n$ (lembrando que de modo geral, n não precisa ser necessariamente primo) é formado pelas classes de congruências de $\mathbb{Z}$ módulo n.

Exemplo 4.2. Vamos determinar o conjunto $\mathbb{Z}_2$. Sabemos que os números inteiros quando divididos por 2, deixam resto 0 ou resto 1. Isto é, qualquer que seja $x \in \mathbb{Z}$ temos $x \equiv 0 \mod 2$ ou $x \equiv 1 \mod 2$. Portanto $\mathbb{Z}_2$ é formado pelas classes $[0]$ e $[1]$, assim $\mathbb{Z}_2 = \{[0], [1]\}$.

Conjuntos quocientes A/B, onde A e B são conjuntos não vazios, são obtidos de certo modo, dividindo os elementos de A por B em subconjuntos disjuntos. Estes conjuntos são formados por classes de equivalência.

O polinômio $f(x)$ citado na lista, desempenhará um papel central se for irredutível. Fazendo uma analogia dos polinômios com os números inteiros, grosso modo, podemos dizer que o polinômio irredutível é semelhante ao número primo, que diferente do número composto, não possui uma forma fatorada. No caso do polinômio irredutí-

vel a semelhança se dá, no sentido de que, o polinômio irredutível não pode ser escrito como produto de polinômios não constantes.

Neste ponto, talvez seja prematuro detalharmos mais sobre o que é o ideal $(f(x))$. Digamos que ele fará o papel de B do conjunto quociente A/B em $F[x]/(f(x))$.

4.2 GRUPOS

Indubitavelmente, *grupos* compõe um dos pilares da álgebra abstrata. São sistemas com uma operação o que facilita a descrição formal.

Definição 4.5. *Grupo*
Diz-se que um conjunto G de elementos, não vazio, forma um grupo se em G está definida uma operação binária, denominada multiplicação e indicada por $\cdot$ tal que:

1. $a, b \in G$ implica que $a \cdot b \in G$ (fechamento).
2. $a, b, c \in G$ implica que $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ (lei associativa).
3. Existe um elemento $e \in G$ tal que $a \cdot e = e \cdot a = a$ para todo $a \in G$ (existência de um elemento unidade em G).
4. Para todo $a \in G$ existe um elemento $a^{-1} \in G$ tal que $a \cdot a^{-1} = a^{-1} \cdot a = e$ (existência de inversos em G).

Vamos definir um conjunto formado por todas as funções bijetivas no próprio conjunto.

Definição 4.6. Se S é um conjunto não vazio, então $A(S)$ é o *conjunto de todas as funções bijetivas* de S em si mesma.

Afirmamos que o conjunto $A(S)$ com a operação de composição é um grupo. Vamos provar, como exemplo, que o conjunto $A(S_3)$ é um grupo, onde $S_3 = \{1, 2, 3\}$ e a multiplicação é a operação de *composição de funções*.

Definição 4.7. Sejam

$$f : \begin{cases} A & \to & B \\ x & \mapsto & f(x) \end{cases} \qquad \text{e} \qquad g : \begin{cases} B & \to & C \\ y & \mapsto & g(y) \end{cases}$$

funções. Chamamos de *composição de funções* entre f e g e denotamos por $f \circ g$ a função h que corresponde a

$$h : \begin{cases} A & \to & C \\ x & \mapsto & h(x) \end{cases},$$

onde, $h(x) = g\left(f(x)\right)$ e denotamos $g\left(f(x)\right)$ por $\left(f \circ g\right)(x)$.

Sejam $\sigma, \tau, \alpha, \beta, \gamma, \iota$, funções de $S \to S$, definidas tais que:

$$\begin{array}{lll} \sigma(1) = 2, & \sigma(2) = 3, & \sigma(3) = 1 \\ \tau(1) = 1, & \tau(2) = 3, & \tau(3) = 2 \\ \alpha(1) = 3, & \alpha(2) = 1, & \alpha(3) = 2 \\ \beta(1) = 2, & \beta(2) = 1, & \beta(3) = 3 \\ \gamma(1) = 3, & \gamma(2) = 2, & \gamma(3) = 1 \\ \iota(1) = 1, & \iota(2) = 2, & \iota(3) = 3. \end{array}$$

Agora que definimos o nosso grupo como sendo $A(S_3) = \{\sigma, \tau, \alpha, \beta, \gamma, \iota\}$. Vamos então analisar os axiomas para verificar que este conjunto se trata de um grupo.

1. **Fechamento**

   É fácil perceber que a composição é fechada para quaisquer que sejam os elementos de $A(S_3)$, ou seja, dado $A(S_3) = \{\sigma, \tau, \alpha, \beta, \gamma, \iota\}$, sabemos que a composição de quaisquer duas funções deste conjunto equivale a uma única função do mesmo conjunto. Façamos um exemplo:

$$(\alpha \circ \beta)(1) = \beta(\alpha(1)) = \beta(3) = 3$$
$$(\alpha \circ \beta)(2) = \beta(\alpha(2)) = \beta(1) = 2$$
$$(\alpha \circ \beta)(3) = \beta(\alpha(3)) = \beta(2) = 1.$$

Notamos assim que $\alpha \circ \beta = \gamma$ e $\gamma \in A(S_3)$. A rigor devemos mostrar para todas as composições possíveis em $A(S_3)$. Com o auxílio da Tabela 1 colocaremos todos estes resultados, entretanto as passagens foram omitidas, pois são imediatas.

$\circ$	σ	τ	α	β	γ	ι
σ	α	γ	ι	τ	β	σ
τ	β	ι	γ	σ	α	τ
α	ι	β	σ	γ	τ	α
β	γ	α	τ	ι	σ	β
γ	τ	σ	β	α	ι	γ
ι	σ	τ	α	β	γ	ι

Tabela 1: A composição é feita na seguinte ordem: compomos a função da linha com a da coluna. Assim, o exemplo dado $\alpha \circ \beta = \gamma$, está na linha 3, coluna 4. Isto contando a partir da parte interna da tabela.

2. **Associativa**

Mostraremos que $(\alpha \circ \beta) \circ \gamma = \alpha \circ (\beta \circ \gamma)$. Observe que:

$$((\alpha \circ \beta) \circ \gamma)(1) = \gamma(\beta(\alpha(1))) = \gamma(\beta(3)) = \gamma(3) = 1 = \alpha(2) = \alpha(\gamma(2)) = \alpha(\gamma(\beta(1))) = (\alpha \circ (\beta \circ \gamma))(1)$$
$$((\alpha \circ \beta) \circ \gamma)(2) = \gamma(\beta(\alpha(2))) = \gamma(\beta(1)) = \gamma(2) = 2 = \alpha(3) = \alpha(\gamma(1)) = \alpha(\gamma(\beta(2))) = (\alpha \circ (\beta \circ \gamma))(2)$$
$$((\alpha \circ \beta) \circ \gamma)(3) = \gamma(\beta(\alpha(3))) = \gamma(\beta(2)) = \gamma(1) = 3 = \alpha(1) = \alpha(\gamma(3)) = \alpha(\gamma(\beta(3))) = (\alpha \circ (\beta \circ \gamma))(3).$$

E portanto $(\alpha \circ \beta) \circ \gamma = \iota = \alpha \circ (\beta \circ \gamma)$.

Assim como no item anterior, a rigor é necessário fazer para todos os casos. Faremos em duas partes. Na primeira parte mostraremos o resultado para a operação feita na seguinte ordem: operamos a primeira função com a segunda e depois o

resultado com a terceira. Na segunda parte vamos operar a segunda função com a terceira e depois a primeira com o resultado obtido da operação anterior.

Vejamos um exemplo, na primeira parte tomando α, β e γ, vamos operar da seguinte maneira:

$$(\alpha \circ \beta) \circ \gamma.$$

Com o auxílio da Tabela 1, temos que $\alpha \circ \beta = \gamma$ e com auxílio da mesma tabela temos que $\gamma \circ \gamma = \iota$, logo $(\alpha \circ \beta) \circ \gamma = \iota$.

Na segunda parte, vamos operar da seguinte maneira:

$$\alpha \circ (\beta \circ \gamma).$$

Do mesmo modo que fizemos na primeira parte, usaremos a Tabela 1 para procedermos com as operações, deste modo temos:

$$\alpha \circ (\beta \circ \gamma) = \alpha \circ \sigma = \iota.$$

E assim mostramos que $(\alpha \circ \beta) \circ \gamma = \iota = \alpha \circ (\beta \circ \gamma)$.

Antes de prosseguirmos iremos excluir da tabela os casos em que a função ι está presente, pois se trata da função identidade do Exemplo 1.1, assim temos para qualquer elemento λ de $A(S)$:

$$\lambda \circ \iota = \iota \circ \lambda = \lambda. \tag{4.1}$$

A validade da equação (4.1) será mostrado no próximo item. Portanto, para quaisquer dois elementos λ_1 e λ_2 de $A(S)$, temos:

$$\begin{aligned}
(\iota \circ \lambda_1) \circ \lambda_2 &= \lambda_1 \circ \lambda_2 = \iota \circ (\lambda_1 \circ \lambda_2) \\
(\lambda_1 \circ \iota) \circ \lambda_2 &= \lambda_1 \circ \lambda_2 = \lambda_1 \circ (\iota \circ \lambda_2) \\
(\lambda_1 \circ \lambda_2) \circ \iota &= \lambda_1 \circ \lambda_2 = \lambda_1 \circ (\lambda_2 \circ \iota).
\end{aligned}$$

A Tabela 2 a seguir será composta pela primeira parte do processo descrito anteriormente.

		Coluna				
		1	2	3	4	5
	$\circ$	σ	τ	α	β	γ
1	$\sigma \circ \sigma = \alpha$	ι	β	σ	γ	τ
2	$\tau \circ \sigma = \beta$	γ	α	τ	ι	σ
3	$\alpha \circ \sigma = \iota$	σ	τ	α	β	γ
4	$\beta \circ \sigma = \gamma$	τ	σ	β	α	ι
5	$\gamma \circ \sigma = \tau$	β	ι	γ	σ	α
6	$\sigma \circ \tau = \gamma$	τ	σ	β	α	ι
7	$\tau \circ \tau = \iota$	σ	τ	α	β	γ
8	$\alpha \circ \tau = \beta$	γ	α	τ	ι	σ
9	$\beta \circ \tau = \alpha$	ι	β	σ	γ	τ
10	$\gamma \circ \tau = \sigma$	α	γ	ι	τ	β
11	$\sigma \circ \alpha = \iota$	σ	τ	α	β	γ
12	$\tau \circ \alpha = \gamma$	τ	σ	β	α	ι
13	$\alpha \circ \alpha = \sigma$	α	γ	ι	τ	β
14	$\beta \circ \alpha = \tau$	β	ι	γ	σ	α
15	$\gamma \circ \alpha = \beta$	γ	α	τ	ι	σ
16	$\sigma \circ \beta = \tau$	β	ι	γ	σ	α
17	$\tau \circ \beta = \sigma$	α	γ	ι	τ	β
18	$\alpha \circ \beta = \gamma$	τ	σ	β	α	ι
19	$\beta \circ \beta = \iota$	σ	τ	α	β	γ
20	$\gamma \circ \beta = \alpha$	ι	β	σ	γ	τ
21	$\sigma \circ \gamma = \beta$	γ	α	τ	ι	σ
22	$\tau \circ \gamma = \tau$	β	ι	γ	σ	α
23	$\alpha \circ \gamma = \tau$	β	ι	γ	σ	α
24	$\beta \circ \gamma = \sigma$	α	γ	ι	τ	β
25	$\gamma \circ \gamma = \iota$	σ	τ	α	β	γ

Linha

Tabela 2: A composição é feita na seguinte ordem: compomos a função da coluna com a da linha. Assim para o exemplo $(\alpha \circ \beta) \circ \gamma$, o resultado está na linha 18, coluna 5. Logo $(\alpha \circ \beta) \circ \gamma = \iota$.

Para fazermos a outra associação da composição, vamos dividir a Tabela 2 em 5 partes. Desta maneira, temos as Tabelas 3, 4, 5, 6 e 7.

Para facilitar a verificação mantivemos a sequência das funções com a mesma numeração de linha.

Todas as tabelas são agrupadas com cinco associações. Fazemos primeiro a composição que está na coluna e cujo o resultado já está expresso na tabela e depois compomos a linha com o resultado já expresso da coluna. Assim como fizemos anteriormente, retiramos os resultados da Tabela 1.

			COLUNA				
			1	2	3	4	5
		$\circ$	$\sigma \circ \sigma = \alpha$	$\sigma \circ \tau = \gamma$	$\sigma \circ \alpha = \iota$	$\sigma \circ \beta = \tau$	$\sigma \circ \gamma = \beta$
LINHA	1	σ	ι	β	σ	γ	τ
	2	τ	γ	σ	β	α	ι
	3	α	σ	τ	α	β	γ
	4	β	τ	ι	γ	σ	α
	5	γ	β	α	τ	ι	σ

Tabela 3: Aqui está a segunda parte das associações feitas da linha 1 à linha 5.

			COLUNA				
			1	2	3	4	5
		$\circ$	$\tau \circ \sigma = \beta$	$\tau \circ \tau = \iota$	$\tau \circ \alpha = \gamma$	$\tau \circ \beta = \sigma$	$\tau \circ \gamma = \alpha$
LINHA	6	σ	τ	σ	β	α	ι
	7	τ	σ	τ	α	β	γ
	8	α	γ	α	τ	ι	σ
	9	β	ι	β	σ	γ	τ
	10	γ	α	γ	ι	τ	β

Tabela 4: Aqui está a segunda parte das associações feitas da linha 6 à linha 10.

		Coluna				
		1	2	3	4	5
	$\circ$	$\alpha \circ \sigma = \iota$	$\alpha \circ \tau = \beta$	$\alpha \circ \alpha = \sigma$	$\alpha \circ \beta = \gamma$	$\alpha \circ \gamma = \tau$
Linha 11	σ	σ	τ	α	β	γ
Linha 12	τ	τ	σ	β	α	ι
Linha 13	α	α	γ	ι	τ	β
Linha 14	β	β	ι	γ	σ	α
Linha 15	γ	γ	α	τ	ι	σ

Tabela 5: Aqui está a segunda parte das associações feitas da linha 11 à linha 15.

Na Tabela 6, está o exemplo citado na Tabela 2. Lá a composição é feita na ordem $(\alpha \circ \beta) \circ \gamma = \iota$, aqui fazemos $\alpha \circ (\beta \circ \gamma)$ que também resulta em ι, cuja a solução está na linha numerada 18 e coluna 5.

		Coluna				
		1	2	3	4	5
	$\circ$	$\beta \circ \sigma = \gamma$	$\beta \circ \tau = \alpha$	$\beta \circ \alpha = \tau$	$\beta \circ \beta = \iota$	$\beta \circ \gamma = \sigma$
Linha 16	σ	β	ι	γ	σ	α
Linha 17	τ	α	γ	ι	τ	β
Linha 18	α	τ	σ	β	α	ι
Linha 19	β	σ	τ	α	β	γ
Linha 20	γ	ι	β	σ	γ	τ

Tabela 6: Aqui está a segunda parte das associações feitas da linha 16 à linha 20.

Note que na Tabela 2, temos $(\alpha \circ \beta) \circ \gamma$. Assim fazemos primeiro $\alpha \circ \beta = \gamma$ e em seguida $\lambda \circ \gamma = \iota$. Enquanto na Tabela 6, temos $\alpha \circ (\beta \circ \gamma)$. Então fazemos primeiro $(\beta \circ \gamma) = \sigma$ e depois $\alpha \circ \sigma = \iota$. Mostrando assim que a associatividade é válida para composição feita pela linha 18 e coluna 5.

Finalmente apresentamos a seguir a última tabela.

			COLUNA				
			1	2	3	4	5
		$\circ$	$\gamma\circ\sigma=\tau$	$\gamma\circ\tau=\sigma$	$\gamma\circ\alpha=\beta$	$\gamma\circ\beta=\alpha$	$\gamma\circ\gamma=\iota$
LINHA	21	σ	γ	α	τ	ι	σ
	22	τ	β	ι	γ	σ	α
	23	α	β	ι	γ	σ	α
	24	β	α	γ	ι	τ	β
	25	γ	σ	τ	α	β	γ

Tabela 7: Aqui está a segunda parte das associações feitas da linha 21 à linha 25.

Construímos as linhas numeradas para corresponder com as associações, conforme explicado no início do item. E como os resultados das linhas correspondentes são todos iguais, temos que para quaisquer λ_1, λ_2 e λ_3 em $A(S)$, segue que $(\lambda_1 \circ \lambda_2) \circ \lambda_3 = \lambda_1 \circ (\lambda_2 \circ \lambda_3)$ e portanto $A(S)$ é associativo.

3. **Existência do elemento neutro**

O elemento neutro se trata da função ι, note que:

$$(\alpha\circ\iota)(1)=\iota(\alpha(1))=\iota(3)=3=\alpha(1)=\alpha(\iota(1))=(\iota\circ\alpha)(1)$$
$$(\alpha\circ\iota)(2)=\iota(\alpha(2))=\iota(1)=1=\alpha(2)=\alpha(\iota(2))=(\iota\circ\alpha)(2)$$
$$(\alpha\circ\iota)(3)=\iota(\alpha(3))=\iota(2)=2=\alpha(3)=\alpha(\iota(3))=(\iota\circ\alpha)(3).$$

Desta maneira, $(\alpha\circ\iota)=(\iota\circ\alpha)=\alpha$. A rigor é necessário mostrar para todas as aplicações de $A(S)$, entretanto, como já dissemos antes, ι se trata da função identidade do Exemplo 1.1, desta maneira, para qualquer função λ em $A(S)$, temos:

$$\iota\circ\lambda=\lambda.$$

Por outro lado se fizermos primeiro a função λ, e em seguida ι esta não irá alterar a função λ, logo $\lambda \circ \iota = \lambda$. E assim:

$$\iota \circ \lambda = \lambda = \lambda \circ \iota.$$

E portanto ι é o elemento neutro de $A(S)$.

4. **Cada elemento do conjunto possui inverso**

   Devemos mostrar que cada elemento do conjunto $A(S)$ possui inverso. Lembrando que demonstramos no item anterior que ι é o elemento neutro de $A(S)$. Nossa tarefa se resume a encontrar para cada elemento λ em $A(S)$, um elemento λ^{-1} em $A(S)$ tal que $\lambda \circ \lambda^{-1} = \lambda^{-1} \circ \lambda = \iota$.

   Mais uma vez usaremos a Tabela 1 para compormos as funções.

   Note que:

$$\sigma \circ \alpha = \alpha \circ \sigma = \iota$$
$$\tau \circ \tau = \tau \circ \tau = \iota$$
$$\beta \circ \beta = \beta \circ \beta = \iota$$
$$\gamma \circ \gamma = \gamma \circ \gamma = \iota$$
$$\iota \circ \iota = \iota \circ \iota = \iota.$$

   E portanto temos que o inverso de σ é α, de τ é τ, de α é σ, de β é β, de γ é γ e como era de se esperar, de ι é ι . Logo cada elemento de $A(S)$ possui inverso.

Assim mostramos que $A(S_3)$ é um grupo.

Como vimos no exemplo ilustrado acima, não há necessidade da comutatividade. Neste caso, temos $\alpha \circ \beta \neq \beta \circ \alpha$. Grupos que são comutativos tem interesse especial no estudo de grupos. Estes grupos são chamados de *abelianos* [1].

Definição 4.8. Um grupo G em que todo $a, b \in G$, $a \cdot b = b \cdot a$ é chamado de *abeliano* (ou comutativo). Um grupo que não é abeliano é simplesmente chamado de *não abeliano*.

1 Nome dado em homenagem ao matemático norueguês *Niels Henrik Abel*, (1802-1829).

Podemos ter um subconjunto H de G, tal que G seja um grupo. Caso H com a operação de G também seja grupo, este conjunto pode ter características importantes. Isso motiva a seguinte definição:

4.2.1 *Subgrupo*

Definição 4.9. Um subconjunto H de um grupo G é dito um subgrupo de G se, com relação ao produto em G, o próprio H forma um grupo.

Uma vez que $H \subset G$ e G é grupo, será que é necessário verificarmos todos os 4 axiomas para nos certificarmos que H é grupo? Como veremos nos dois lemas a seguir, é possível decidir se H é grupo ou não por outros meios.

Por simplicidade de notação eliminaremos, de agora em diante, o ponto em $a \cdot b$ e indicaremos este produto simplesmente por ab.

Lema 4.10. *Um subconjunto não vazio H do grupo G é um subgrupo de G se, e somente se,*

1. *$a, b \in H$ implica que $ab \in H$.*
2. *$a \in H$ implica que $a^{-1} \in H$.*

Demonstração. Se H é um subgrupo de G evidentemente 1 e 2 são válidos. Reciprocamente, para mostrar que H é grupo, basta mostrar que $e \in H$, uma vez que a lei associativa é claramente válida por $H \subset G$ e G ser grupo. Para isto, observamos que se $a \in H$, então por 2, $a^{-1} \in H$ e portanto, por 1, temos $aa^{-1} = e \in H$, o que conclui a demonstração. □

A Definição 4.2 de congruência modular que foi dada no início do capítulo, tem como operação a adição, mas em grupos somente definimos a operação de multiplicação. Note que dados a e b, verificamos se m divide a adição de a pelo oposto (ou simétrico aditivo) de b, ou seja $-b$, assim verificamos se $m \mid a + (-b)$. Baseado nesta ideia de congruência, tendo em vista o grupo H, dados dois elementos $a, b \in H$, faremos a operação de a com o simétrico de b, mas como se trata da operação de multiplicação,

efetuamos a multiplicação de a pelo inverso multiplicativo de b que é b^{-1}, ou seja, fazemos ab^{-1}. Com isto, podemos enunciar a seguinte definição:

Definição 4.11. Sejam G um grupo e H um subgrupo de G. Para $a, b \in G$ dizemos que a é congruente a b módulo H, indicado por $a \equiv b \mod H$, se $ab^{-1} \in H$.

Agora tomando a Definição 4.1 de equivalência e a Definição 4.2 de congruência como partida, chegamos ao Lema 4.12, que está demonstrado em [6].

Lema 4.12. *A relação $a \equiv b \mod H$ é uma relação de equivalência.*

Ou seja, temos as seguintes relações:

1. $a \equiv a$;
2. $a \equiv b$ implica $b \equiv a$;
3. $a \equiv b$ e $b \equiv c$ então $a \equiv c$.

E com isto chegamos a Definição 4.13 de *classes laterais*, que é similar a Definição 4.3 de classe de congruência.

Definição 4.13. Se H é um subgrupo de G, $a \in G$, então $Ha = \{ha; h \in H\}$. Ha é denominada *classe lateral à direita de H em G*.

Observação 4.14. A Definição 4.11 nos conduz a Definição 4.13 como sendo de classe lateral à direita. Lembrando que a comutatividade não é uma exigência para grupos, note que justificamos acima $a \equiv b \mod H$, operando com o inverso multiplicativo de b, pela direita de a. Assim temos:

$$a \equiv b \mod H \iff ab^{-1} \equiv bb^{-1} = e \mod H.$$

Entretanto, poderíamos ter optado em multiplicar pelo inverso multiplicativo de b pela esquerda de a,

$$a \equiv b \mod H \iff b^{-1}a \equiv b^{-1}b = e \mod H,$$

desta maneira, a Definição 4.11 ficaria:

Definição 4.15. Sejam G um grupo e H um subgrupo de G. Para $a, b \in G$ dizemos que a é congruente a $b \mod H$, indicado por $a \equiv b \mod H$, se $b^{-1}a \in H$.

Assim seríamos conduzidos a definir classe lateral à esquerda, como:

Definição 4.16. Se H é um subgrupo de G, $a \in G$, então $aH = \{ah; h \in H\}$. Denominamos de *aH classe lateral à esquerda de H em G.*

Note que se considerarmos a Definição 4.16, teremos $[a] = aH$.

Lema 4.17. *Para todo $a \in G$,*

$$Ha = \{x \in G; a \equiv x \mod H\}.$$

Demonstração. Seja $[a] = \{x \in G; a \equiv x \mod H\}$. Em primeiro lugar, mostraremos que $Ha \subset [a]$. De fato, se $h \in H$, então $a(ha)^{-1} = a\left(h^{-1}a^{-1}\right) = h^{-1}$, que está em H, pois H é um subgrupo de G. Pela definição de congruência $\mod H$ isto implica que $ha \in [a]$ para todo $h \in H$, e então $Ha \subset [a]$.

Suponhamos, agora que $x \in [a]$. Assim $ax^{-1} \in H$, então $\left(ax^{-1}\right)^{-1} = xa^{-1}$ também está em H. Isto é, $xa^{-1} = h$ para algum $h \in H$. Multiplicando ambos os membros por a pela direita, chegamos a $x = ha$ e então $x \in Ha$. Assim $Ha \supset [a]$. Tendo demonstrado as duas inclusões $Ha \subset [a]$ e $Ha \supset [a]$, podemos concluir que $a = [Ha]$, que é a asserção do lema. [6] □

Podemos agora dizer que Ha é uma classe de equivalência de a em G. Associando esta informação ao Teorema 4.4 da teoria de conjuntos, chegamos a conclusão que quaisquer classes laterais à direita de H em G são idênticas ou não possuem elementos em comum.

Agora podemos definir entre as classes $[a] = \{x \in G; a \equiv x \mod H\}$ e $[b] = \{y \in G; b \equiv y \mod H\}$, a seguinte operação:

$$[a]\,[b] = [ab].$$

Notemos que:

$$a \equiv x \mod H \tag{4.2}$$
$$b \equiv y \mod H. \tag{4.3}$$

Multiplicando as relações de congruências 4.2 e 4.3 membro a membro, temos:

$$ab \equiv xy \mod H.$$

Isto mostra que $[a][b] = [ab]$, está bem definida, pois $[ab] = \{xy \in G; ab \equiv xy \mod H\}$. Uma vez que $ax^{-1}by^{-1} \equiv xx^{-1}yy^{-1} = e \in H$.

Outro lema importante, demonstrado em [6], faz relação entre duas classes laterais.

Lema 4.18. *Existe uma correspondência biunívoca entre duas quaisquer classes laterais à direita de H em G.*

O Lema 4.18 é de especial interesse no caso de H ser um grupo finito, pois isto indica que o número de elementos de duas classes laterais à direita são iguais.

No Capítulo 1, a Definição 1.3 diz que a cardinalidade de um conjunto é a quantidade de elementos do conjunto. De maneira análoga, definiremos agora a *ordem* de um grupo.

Definição 4.19. Definimos a *ordem* de um grupo G, como sendo o número cardinal n do conjunto G e denotamos por $o(G) = n$.

Vamos apresentar um importante teorema devido a Lagrange [2], que relaciona a ordem $o(H)$ de uma classe lateral de um subgrupo H de G com a ordem $o(G)$ da classe lateral de G.

Teorema 4.20. *de Lagrange*
Se H é um subgrupo de um grupo finito de G, então $o(H)$ é divisor de $o(G)$.

2 Joseph Louis Lagrange foi um matemático italiano que nasceu em *Turim*, no dia 25 de janeiro de 1736 e faleceu em *Paris*, no dia 10 de abril de 1813

Demonstração. Suponhamos que $o(H) = n$ seja quantidade de classes laterais a direita de H em G. Como G é um grupo finito que contém H, pelos Lemas 4.17 e 4.18 duas quaisquer classes laterais à direita de H em G não têm elemento comum, e cada uma tem $o(H) = n$ elementos. Por outro lado, como cada $a \in G$ está em uma única classe lateral à direita, Ha, implica que todas as classes laterais à direita de G, seja uma quantidade k inteira de classes laterais de $o(H)$, ou seja, $o(G) = kn$. Logo $o(G) = ko(H)$ e portanto $o(H)$ é divisor de $o(G)$. □

Definição 4.21. Se H é um subgrupo de G, o *índice de H em G* é o número de classes laterais à direita de H em G. Denotaremos por $i_G(H)$.

Pelo Teorema 4.20 de Lagrange, podemos escrever para o caso de G finito,

$$i_G(H) = \frac{o(G)}{o(H)}.$$

Definição 4.22. Se G é um grupo e $a \in G$, a *ordem de a*, denotada por $o(a)$, é o menor inteiro positivo m tal que $a^m = e$. Caso não exista m que satisfaça a igualdade dizemos que a tem *ordem infinita*.

Vou enunciar dois corolários que são obtidos do Teorema 4.20 de Lagrange, cujas as demonstrações estão em [6].

Corolário 4.23. *Se G é um grupo finito e $a \in G$, então $o(a) \mid o(G)$.*

Corolário 4.24. *Se G é um grupo finito e $a \in G$, então $a^{o(G)} = e$.*

4.2.2 *Subgrupos Normais e Grupos Quocientes*

Agora vamos construir subgrupos de modo bem especial. De modo que as classes laterais à esquerda e à direita coincidam.

Definição 4.25. Um subgrupo N de G é dito um *subgrupo normal* de G se para todo $g \in G$ e $n \in N$, $gng^{-1} \in N$.

Como dissemos acima, nossa intenção era construir um subgrupo de G tal que as classes laterais à esquerda e à direita coincidam, ou seja, $gN = Ng$. E isto foi feito quando definimos um subgrupo N de G como sendo um subgrupo normal de G se para todo $g \in G$ e $n \in N$, $gng^{-1} \in N$, pois supondo $n_1, n_2 \in N$ a definição nos diz que $gn_1g^{-1} = n_2$, desta forma temos

$$gn_1g^{-1} = n_2 \iff gn_1g^{-1}g = n_2g \iff \underbrace{gn_1}_{\in gN} = \underbrace{n_2g}_{\in Ng}.$$

Enunciaremos três lemas que estão demonstrados em [6], que nos permitirá construir um grupo a partir de uma classe lateral.

Lema 4.26. *N é um subgrupo normal de G se, e somente se, $gNg^{-1} = N$ para todo $g \in G$.*

Lema 4.27. *O subgrupo N de G é um subgrupo normal de G se, e somente se, toda classe lateral à esquerda de N em G é uma classe lateral à direita de N em G.*

Lema 4.28. *Um subgrupo N de G é um subgrupo normal de G se, e somente se, o produto de duas classes laterais à direita de N em G também é uma classe lateral à direita de N em G.*

Os lemas nos levam a definir uma estrutura quociente, que é de alta relevância na matemática. Note que supondo N um subgrupo normal de G. A fórmula $NaNb = Nab$, para $a, b \in G$ nos remete a construção de um grupo que use tal produto. Assim indicamos por G/N a coleção das classes laterais à direita de N em G (isto é, os elementos de G/N são certos subconjuntos de G) e usamos o produto de subconjuntos de G para obtermos um produto em G/N.

Para este produto afirmamos:

1. $X, Y \in G/N$ implica $XY \in G/N$; pois $X = Na$, $Y = Nb$ para certos $a, b \in G$, e $XY = NaNb = Nab \in G/N$.
2. $X, Y, Z \in G/N$, então $X = Na$, $Y = Nb$, $Z = Nc$ com $a, b, c \in G$ e então $(XY)Z = (NaNb)Nc = N(ab)Nc = N(ab)c = Na(bc)$ (pois G é associativo) $= Na(Nbc) = Na(NbNc) = X(YZ)$, Assim o produto em G/N satisfaz a lei associativa.
3. Consideremos o elemento $N = Ne \in G/N$. Se $X \in G/N$ e $X = Na$ com $a \in G$, então $XN = NaNe = Nae = Na = X$, e analogamente $NX = X$. Consequentemente, Ne é um elemento unidade para G/N.

4. Suponhamos $X = Na \in G/N$ (onde $a \in G$); assim $Na^{-1} \in G/N$, e $NaNa^{-1} = Naa^{-1} = Ne$. Analogamente, $Na^{-1}Na = Ne$. Portanto, Na^{-1} é o inverso de Na em G/N.

Mas um sistema que satisfaz $1, 2, 3, 4$ é exatamente o que denominamos um grupo. Isto é:

Teorema 4.29. *Se G é um grupo, N um subgrupo normal de G, então G/N também é um grupo. É denominado o grupo quociente ou o grupo fator G por N.*

Lema 4.30. *Se G é um grupo finito e N é subgrupo normal de G, então*

$$o(G/N) = \frac{o(G)}{o(N)}.$$

Demonstração. Como N é um subgrupo normal de G, temos que G/N é formado por $i_G(N) = \frac{o(G)}{o(N)}$ elementos, que são as classes laterais à direita de N em G. □

4.2.3 *Homomorfismo*

Uma função de um sistema algébrico em outro sistema algébrico semelhante que conserve sua estrutura, nos permitirá produzir resultados valiosos para a álgebra. Esta função é chamada de *homomorfismo*, definiremos a seguir homomorfismo para grupos.

Definição 4.31. Uma função ϕ de um grupo G em um grupo $\overline{G}$ é dita um *homomorfismo* se para todos $a, b \in G$, $\phi(ab) = \phi(a)\phi(b)$.

Observamos que no primeiro membro desta igualdade, isto é, no termo $\phi(ab)$ o produto ab é feito em G usando o produto de elementos de G enquanto que no segundo membro, ou seja, no termo $\phi(a)\phi(b)$, o produto é feito entre elementos de $\overline{G}$.

O próximo lema nos mostra que um grupo G e um grupo quociente G/N, onde N é um subgrupo normal de G, têm mesma estrutura. Assim existe um homomorfismo entre G/N e G.

Lema 4.32. *Suponhamos que G seja um grupo e N um subgrupo normal de G; definamos a função de G em G/N por $\phi(x) = Nx$ para todo $x \in G$. Então ϕ é um* homomorfismo sobrejetivo *de G em G/N.*

Demonstração. Que é sobrejetiva é trivial pois todo elemento $X \in G/N$ é da forma $X = Ny$, $y \in G$, portanto $X = \phi(y)$. Para verificar a propriedade multiplicativa requerida para ϕ ser homomorfismo basta notar que se $x, y \in G$, é $\phi(xy) = Nxy = NxNy = \phi(x)\phi(y)$. □

Mostramos acima um homomorfismo sobrejetivo, mas e quanto a *homomorfismos injetivos*? A definição a seguir nos mostra que o homomorfismo injetivo é uma condição bem mais restrita.

Definição 4.33. Se ϕ é um homomorfismo de G em $\overline{G}$, o *núcleo* (ou *kernel*) de ϕ, K_ϕ, é definido por $K_\phi = \{x \in G;\ \phi(x) = \overline{e}\}$, onde $\overline{e}$ é o elemento unidade de $\overline{G}$.

A definição nos mostra que todo elemento x em G que pertence ao núcleo é levado à $\overline{e}$ em $\overline{G}$, ou seja, se o núcleo não for um conjunto unitário não haverá a injetividade.

Outro aspecto que temos que mostrar é que K_ϕ não é vazio. Faremos isto no lema a seguir.

Lema 4.34. *Se ϕ é um homomorfismo de G em $\overline{G}$, então:*

1. *$\phi(e) = \overline{e}$, o elemento unidade de $\overline{G}$.*
2. *$\phi(x^{-1}) = \phi(x)^{-1}$ para todo $x \in G$.*

Consideremos agora um homomorfismo sobrejetivo do grupo G em $\overline{G}$, e suponhamos que K seja o núcleo de ϕ. Tomando um elemento $x \in G$ tal que $\phi(x) = \overline{g}$, com $\overline{g} \in \overline{G}$, dizemos que x é uma *imagem inversa de* $\overline{g}$. O lema a seguir nos diz sobre a forma das imagens inversas de x.

Lema 4.35. *Se ϕ é um homomorfismo sobrejetivo de G em $\overline{G}$ com núcleo K, então o conjunto das imagens inversas de $\overline{g} \in \overline{G}$ com relação a ϕ é dada por Kx, onde x é uma imagem inversa particular de $\overline{g}$ em G.*

Uma situação específica de $K = \{e\}$, pelo Lema 4.35 nos leva a concluir que ϕ é uma função injetiva, pois qualquer $\overline{g} \in \overline{G}$, tem exatamente uma imagem inversa e reciprocamente se ϕ é um homomorfismo injetivo de G em $\overline{G}$ seu núcleo consiste exatamente em e.

Definição 4.36. Um homomorfismo ϕ de G em $\overline{G}$ é dito um *monomorfismo* se ϕ é injetiva.

O próximo passo agora é definir uma função que seja injetiva e sobrejetiva.

Definição 4.37. Dois grupos G e G^* são ditos *isomorfos* se existe um homomorfismo bijetivo (ou seja, um isomorfismo) de G em G^*. Neste caso escrevemos $G \approx G^*$.

Esta é claramente uma relação de equivalência, ou seja, temos:

1. $G \approx G$.
2. $G \approx G^*$ implica $G^* \approx G$.
3. $G \approx G^*$ e $G^* \approx G^{**}$ implicam $G \approx G^{**}$.

Quando dois grupos são isomorfos, então, de certo modo, eles são iguais. Eles diferem quanto a natureza dos conjuntos. O isomorfismo nos dá uma maneira de relacionar os elementos de naturezas diferentes, mas que têm comportamento análogo. Podemos imaginar uma música sendo tocada num violão e que para executá-la usamos uma tablatura. Podemos tocar a mesma música no piano, executar as mesmas notas, entretanto se trata de um instrumento completamente diferente, então uma tablatura feita para violão não nos dará informação para tocar o piano. Entretanto, se tivermos uma partitura que contenha as mesmas informações da tablatura, ela nos permitirá executar a mesma música no piano. O isomorfismo é o mecanismo de tradução entre partitura e a tablatura, que nos permite traduzir a tablatura do violão para executar a música no piano.

Voltando ao Lema 4.35 por um momento, vemos nele um meio de caracterizar em termos do núcleo quando um homomorfismo é realmente um monomorfismo.

Corolário 4.38. *Um homomorfismo ϕ de G em $\overline{G}$ com núcleo K_ϕ é um monomorfismo de G em $\overline{G}$ se, e somente se, $K_\phi = \{e\}$.*

Este corolário nos fornece um método para determinar se dois grupos são isomorfos. Primeiro encontramos um homomorfismo sobrejetivo de um no outro, e então mostra-

mos que o núcleo deste homomorfismo consiste apenas do elemento unidade. Isto é descrito no teorema a seguir.

Teorema 4.39. *Seja ϕ um homomorfismo sobrejetivo de G em $\overline{G}$ com núcleo K. Então $G/K \approx \overline{G}$.*

Um caso particular de homomorfismo é quando existe um isomorfismo de um grupo em si mesmo.

Definição 4.40. Uma função ϕ de um grupo G em si mesmo é um *automorfismo*, se ϕ for um isomorfismo.

4.2.4 *Grupos de Permutação*

Antes de terminar esta seção vamos definir um grupo que será de grande importância para o Capítulo 6, o *grupo simétrico*. Sua importância pode ser ilustrada pelo Teorema de Cayley, cuja a demonstração pode ser encontrada em [6].

Teorema 4.41. *Cayley*
Todo grupo é isomorfo a um subgrupo de $A(S)$ para um S conveniente.

Definição 4.42. Definimos o *grupo simétrico de grau n*, que denotamos por S_n, o grupo de todas as permutações de um conjunto S, com n elementos, munido da operação de composição de funções.

A fim de facilitar a notação, utilizaremos uma notação matricial. Por exemplo, para uma função $\psi : S \to S$ tal que $S = \{1,2,3,4\}$, onde $\psi(1) = 2$, $\psi(2) = 4$, $\psi(3) = 1$ e $\psi(4) = 3$. Escreveremos:

$$\psi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix}.$$

Aqui trataremos de funções de S em S onde $S = \{1,2,\ldots,n\}$, assim a primeira linha será composta pelo domínio, e linha debaixo pela imagem do elemento acima dele. Por isto, olhando para a matriz é fácil perceber que $\psi(3) = 1$. Vejamos outro exemplo:

$$\theta = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 5 & 2 & 6 & 7 & 1 & 4 \end{pmatrix}.$$

Então, neste caso $\theta(1) = 3$, $\theta(2) = 5$, $\theta(3) = 2$, $\theta(4) = 6$, $\theta(5) = 7$, $\theta(6) = 1$ e $\theta(7) = 4$.

Também será de grande utilidade definir uma notação com expoentes para composição de funções. Assim:

Definição 4.43. Dada uma função $\alpha : A \to A$ denotaremos por α^n com $n \in \mathbb{N}$ e $n \geq 2$ a composição $\underbrace{\alpha \circ \alpha \circ \cdots \circ \alpha}_{n \text{ vezes}}$. E definimos $\alpha^1 = \alpha$ e $\alpha^0 = \iota$, onde ι é a função identidade.

Também podemos, elevar uma função a uma potência negativa. Desta forma definimos α^{-n} como sendo a composição de $\underbrace{\alpha^{-1} \circ \alpha^{-1} \circ \cdots \circ \alpha^{-1}}_{n \text{ vezes}}$ com $n \in \mathbb{N}$ e $n \geq 2$.

Outra função que também vamos destacar a notação, é a função inversa de α que é denotada por α^{-1}. Lembrando que $\alpha \circ \alpha^{-1} = \alpha^{-1} \circ \alpha = \iota$. Com isto, para $\theta : S \to S$ com $S = \{1, 2, \ldots, n\}$, onde $\theta \in S_n$ e $r, s \in \mathbb{Z}$, são válidas as seguintes propriedades:

- $\theta^r \circ \theta^s = \theta^{r+s}$.
- $(\theta^r)^s = \theta^{rs}$.
- $\theta^{-r} = \left(\theta^{-1}\right)^r$.
- $\theta^{-1} \circ \theta^1 = \theta^{-1+1} = \theta^0 = \iota$.

Desta maneira, podemos manipular com fórmula, que contenha função que possua como expoente números inteiros. E será isso que acontecerá em nossa próxima definição. Mais uma vez, iremos simplificar a notação. Até o final desta seção suprimiremos o símbolo $\circ$ na composição das funções, assim em vez de escrever $\psi \circ \theta$ escreverei simplesmente $\psi\theta$.

Voltando a matriz de permutação, podemos deixar a notação ainda mais concisa. Para isto devemos definir mais alguns conceitos. Começamos definindo uma relação de equivalência sobre S.

Definição 4.44. Sejam S um conjunto e $\theta \in A(S)$. Então para quaisquer $a, b \in S$ temos $a \equiv_\theta b$ se, e somente se, $b = \theta^i(a)$ para algum $i \in \mathbb{Z}$.

Vamos mostrar que esta relação cumpre os três axiomas de uma relação de equivalência.

1. $a \equiv_\theta a$, pois $a = \theta^0(a) = \iota(a) = a$. (Lembrando que $\theta\theta^{-1} = \theta^0 = \iota$).
2. Se $a \equiv_\theta b$, então $a = \theta^i(b) \Leftrightarrow \theta^{-i}(a) = \theta^{-i}\theta^i(b) \Leftrightarrow \theta^{-i}(a) = b \Leftrightarrow b = \theta^{-i}(a)$, logo $b \equiv_\theta a$.
3. Se $a \equiv_\theta b$, então $a = \theta^i(b)$. Por outro lado se, $b \equiv_\theta c$, então $b = \theta^j(c)$. Assim usando o fato de que $b = \theta^j(c)$ em $a = \theta^i(b)$, obtemos $a = \theta^i(b) = \theta^i\left(\theta^j(c)\right) = \left(\theta^i\theta^j\right)(c) = \theta^{i+j}(c)$. Portanto, $a \equiv_\theta c$.

Esta relação de equivalência, pelo Teorema 4.4 induz uma decomposição de S em subconjuntos disjuntos, que são as classes de equivalência. Estas classes de equivalência de elementos s em S são denominadas *órbitas* de S sob θ. Desta maneira, a órbita de S sob θ consiste de todos os elementos $\theta^i(s)$ com $i \in \mathbb{Z}$.

Vamos ilustrar o que dissemos anteriormente, por meio de um exemplo.

Exemplo 4.3. Considere $S = \{1, 2, \ldots, 6\}$ e $\theta = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 3 & 5 & 6 & 4 \end{pmatrix}$. Queremos aqui determinar os elementos das órbitas de S por θ. Sabemos que estes elementos consistem em $\theta^i(s)$ com $i \in \mathbb{Z}$, assim para $s = 1$, temos:

$$\theta^0(1) = \iota(1) = 1,\ \theta^1(1) = \theta(1) = 2,\ \theta^2(1) = \theta\left(\theta(1)\right) = \theta(2) = 1.$$

e portanto para órbita $s = 1$ por θ os elementos são 1 e 2. Com isso temos $1 \equiv_\theta 2$. Para $s = 3$, temos $\theta^i(3) = 3$ para qualquer i em $\mathbb{Z}$ e portanto 3 é o único elemento para órbita $s = 3$ e assim só vale a congruência trivial $3 \equiv_\theta 3$. E para $s = 4$, temos:

$$\theta(4) = 5,\ \theta^2(4) = \theta\left(\theta(4)\right) = \theta(5) = 6,\ \theta^3(4) = \theta\theta^2(4) = \theta(6) = 4.$$

E desta maneira, chegamos aos elementos $4, 5$ e 6 para órbita de $s = 4$ e com isto temos $4 \equiv_\theta 5 \equiv_\theta 6$.

Determinado os elementos da órbita de s, vamos organizá-los de maneira consistente, para o caso particular de S ser finito. Nesta situação vamos tomar os elementos da órbita, agrupando-os em ênuplas ordenadas, que são denominadas de *ciclos* de θ . Todos os ciclos de θ determinam θ.

Observe que existe um menor inteiro positivo, que denotaremos por l tal que $\theta^l(s) = s$. A órbita de s sob θ é $s, \theta(s), \theta^2(s), \ldots, \theta^{l-1}(s)$. Assim o ciclo para alguma órbita s

por θ é $\left(s, \theta(s), \theta^2(s), \ldots, \theta^{l-1}(s)\right)$. Fizemos isso, a fim de que, pudéssemos visualizar a matriz por intermédios dos ciclos. Isto ficará mais claro no exemplo.

Ilustrando o que está escrito através do Exemplo 4.3, temos para $s = 1$, $\theta(1) = 2$, $\theta^2(1) = \theta(2) = 1$. Assim para $s = 1$ o menor l é igual a 2 e deste modo o ciclo é $(s, \theta(s)) = (1, \theta(1)) = (1, 2)$. Para $s = 3$ temos $\theta(3) = 3$ e portanto $l = 1$ e assim o ciclo para $s = 3$ é $(s) = (3)$. E por fim, para $s = 4$ temos $s(4) = 5$, $s^2(4) = s(5) = 6$ e $s^3(4) = s(6) = 4$. Então temos $l = 3$ e desta maneira $\left(s, \theta(s), \theta^2(s)\right) = \left(4, \theta(4), \theta^2(4)\right) = (4, 5, 6)$.

Dissemos que os ciclos nos permite a visualização da matriz. Note que, no primeiro ciclo temos $(1, 2)$. Assim, o ciclo indica que $\theta(1) = 2$ e $\theta(2) = 1$, ou seja, $\theta : 1 \mapsto 2$ e $\theta : 2 \mapsto 1$. Escrevendo de forma mais livre, temos $1 \to 2 \to 1$, que fecha o ciclo. O três está fixo e para $s = 4$, temos $4 \to 5 \to 6 \to 4$.

Exemplo 4.4. Considere a permutação

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 1 & 6 & 7 & 5 & 2 & 8 & 4 \end{pmatrix}.$$

Vamos escrevê-la como ciclo, conforme fizemos anteriormente. Notamos que $1 \to 3 \to 6 \to 2 \to 1$, terminando assim o ciclo $(1, 3, 6, 2)$. Nesta matriz também podemos obter o ciclo $(4, 7, 8)$, pois $4 \to 7 \to 8 \to 4$. E terminamos percebendo que 5 está fixo, assim podemos representar a matriz de permutação através dos ciclos $(1, 3, 6), (4, 7, 8), (5)$.

Antes de definir o produto entre dois ciclos. Vamos fazer o produto entre duas permutações, usando a representação matricial. A ideia é semelhante a composição entre duas funções. Ao fazer o produto de $\tau = \begin{pmatrix} 1 & \ldots & n \\ i_1 & \ldots & i_n \end{pmatrix}$ por $\psi = \begin{pmatrix} 1 & \ldots & n \\ j_1 & \ldots & j_n \end{pmatrix}$, levamos 1 em i_1 por τ e depois pegamos i_1 e levamos ao j_s correspondente por ψ. Repetimos este procedimento para todos os n elementos de τ, determinando assim $\tau\psi$. Façamos um exemplo:

Exemplo 4.5. Efetuemos o produto $\tau\psi$, onde $\tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix}$ e $\psi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{pmatrix}$. Conforme dito anteriormente, τ leva 1 em 3 e ψ leva 3 em 2, assim $\tau\psi$ leva 1 em 2. Depois fazemos o mesmo para 2 e assim por diante. Deste modo, temos:

$$\tau\psi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 3 & 4 \end{pmatrix}.$$

Vamos agora definir uma multiplicação entre ciclos. Para isto transformaremos os ciclos em matrizes de permutações. Considere o ciclo $(i_1, i_2, \ldots, i_r)$ e ψ a matriz de permutação, tal que ψ leva i_1 em i_2, i_2 em i_3, ..., k_{r-1} em i_r e i_r em i_1 e fixamos todos os outros elementos de S fixos. Façamos um exemplo, se considerarmos o ciclo $(1,4,2,7,3)$ de uma permutação para $S = \{1,2,\ldots,8\}$ o ciclo será representado pela permutação:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 4 & 7 & 1 & 2 & 5 & 6 & 3 & 8 \end{pmatrix}.$$

A multiplicação entre dois ciclos será feita multiplicando as permutações que os ciclos representam.

Exemplo 4.6. Dados $(1,2,3)$ e $(5,6,4,1,8)$ dois ciclos para S com 8 elementos. Efetuemos seu produto.

$$\begin{aligned}(1,2,3)(5,6,4,1,8) &= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 3 & 1 & 4 & 5 & 6 & 7 & 8 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 8 & 2 & 3 & 1 & 6 & 4 & 7 & 5 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 3 & 8 & 1 & 6 & 4 & 7 & 5 \end{pmatrix}.\end{aligned}$$

Podemos observar que os ciclos da permutação

$$\theta = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 2 & 3 & 8 & 1 & 6 & 4 & 7 & 5 & 9 \end{pmatrix}.$$

são $(1,2,3,8,5,6,4)$,(7),(9). E que ao multiplicar estes ciclos obtemos θ. Isto ocorreu não por acaso e é o que mostraremos no lema a seguir.

Lema 4.45. *Toda permutação é produto de seus ciclos.*

Demonstração. Considere uma permutação θ. Assim $\left(s, \theta(s), \theta^2(s), \ldots, \theta^{l-1}(s)\right)$ representam os ciclos de θ. Pela definição da multiplicação de ciclos, temos que os ciclos de θ são disjuntos. A imagem de $s' \in S$ sob θ que é $\theta(s')$, é a mesma que a imagem de s' sob o produto ψ de todos ciclos distintos de θ. Portanto, θ e ψ, têm o mesmo efeito em cada elemento de S, logo $\theta = \psi$ o que conclui a demonstração. □

O Lema 4.46 também pode ser enunciado da seguinte maneira: *"toda permutação pode ser expressa de maneira única como um produto de ciclos disjuntos"*.

Consideremos agora a possibilidade de escrever os ciclos como um produto de outros ciclos. Dado o m-ciclo $(1, 2, \ldots, m)$, pelo que já definimos, podemos decompô-lo, tal que $(1, 2, \ldots, m) = (1, 2)(1, 3)\ldots(1, m)$. No entanto esta decomposição não é única. Por exemplo, $(1, 2, 3) = (1, 2)(1, 3) = (3, 1)(3, 2)$.

Lema 4.46. *Toda permutação é um produto de 2-ciclo.*

Demonstração. Seja θ uma permutação. Pelo Lema 4.45, θ pode ser escrita como um produto de seus ciclos. Desta forma podemos escrever:

$$\theta = (a_1, a_2, \ldots, a_n) = (a_{i_1}, \ldots, a_{i_r})\ldots(a_{j_1}, \ldots, a_{j_s}).$$

Por sua vez, cada ciclo de θ é da forma $(a_{i_1}, a_{i_2}, \ldots, a_{i_m}) = (a_{i_1}, a_{i_2})\ldots(a_{i_1}, a_{i_m})$. Logo

$$\begin{aligned}\theta = (a_1, a_2, \ldots, a_n) &= (a_{i_1}, \ldots, a_{i_r})\ldots(a_{j_1}, \ldots, a_{j_s}) \\ &= (a_{i_1}, a_{i_2})\ldots(a_{i_1}a_{i_r})\ldots(a_{j_1}, a_{j_2})\ldots(a_{j_1}a_{j_s}).\end{aligned}$$

□

4.3 ANEL

Nesta seção avançaremos nossos estudos, vamos agora definir outra estrutura que juntamente com *grupos* constitui um dos pilares da álgebra moderna, estamos falando de *anéis*.

Definição 4.47. Um conjunto não vazio R é dito um anel associativo se em R estão definidas duas operações, indicadas por $+$ e $\cdot$ respectivamente, tais que para todos a, b e c em R:

1. $a + b$ está em R.
2. $a + b = b + a$.
3. $(a + b) + c = a + (b + c)$.
4. Existe um elemento 0 em R tal que $a + 0 = a$ (para cada a em R).
5. Existe um elemento $-a$ em R tal que $a + (-a) = 0$.
6. $a \cdot b$ está em R.

7. $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.

8. $a \cdot (b + c) = a \cdot b + a \cdot c$ e $(b + c) \cdot a = b \cdot a + c \cdot a$ (as duas leis distributivas).

Podemos notar que os axiomas de 1 a 5 nos revelam que R é um grupo abeliano com relação à operação +, usualmente definida de adição. Os axiomas 6 e 7 diz que R é fechado em relação a uma operação, $\cdot$, denominada multiplicação. E finalmente o axioma 8 une as duas operações em R.

No axioma 4 o 0 é denominado elemento *nulo*, enquanto no axioma 5 o $-a$ é usualmente chamado de *oposto* de a, contudo como observamos acima, este axioma também é um axioma de grupo e assim $-a$ também é por vezes chamado de *inverso aditivo*, uma vez que R é um grupo abeliano em relação à operação de adição.

Analogamente temos que o elemento 0 faz às vezes de e, ou seja, $e = 0$ ao considerarmos o conjunto R um grupo em relação a operação de adição. Contudo, temos que tomar **cuidado!** Dentro dos anéis existem duas operações, assim reservamos nomes distintos como fizemos com os inversos, deste modo o elemento unidade do conjunto não é o 0, este é o elemento nulo, enquanto e continua sendo o elemento unidade e está relacionado com a operação de multiplicação.

Vamos agora definir algumas classes especiais de anéis.

Por simplicidade de notação eliminaremos, de agora em diante, o ponto em $a \cdot b$ e indicaremos este produto simplesmente por ab.

Definição 4.48. Se R é um anel comutativo, então $a \neq 0 \in R$ é dito um *divisor do zero* se existe um $b \in R$, $b \neq 0$, tal que $ab = 0$.

Definição 4.49. Um anel comutativo é um *anel de integridade* se não possui divisores do zero.

O anel dos inteiros é, naturalmente, um exemplo de anel de integridade.

Definição 4.50. Um anel é dito anel com divisão se seus elementos não nulos formam um grupo com relação à multiplicação.

O elemento unidade com relação à multiplicação será indicado por 1, e o inverso de um elemento a com relação à multiplicação será indicado por a^{-1}.

Finalmente, introduzimos a definição da estrutura algébrica de muita importância para a álgebra, o corpo.

Definição 4.51. Um *corpo* é um anel com divisão comutativo.

Lema 4.52. *Se R é um anel, então, para todos $a, b \in R$,*

1. $a0 = 0a = 0$.
2. $a(-b) = (-a)b = -(ab)$.
3. $(-a)(-b) = ab$.

Se, além disso, R possui um elemento unidade 1*, então*

4. $(-1)a = -a$.
5. $(-1)(-1) = 1$.

Demonstração.

1. Se $a \in R$ então $a0 = a(0+0)$. Como R é um anel, podemos aplicar a lei distributiva (axioma 8) e usar o fato de todos os elementos possuírem opostos (axioma 5), ou seja, podemos afirmar que $a0 + (-a0) = 0$, obtemos:

$$\begin{aligned} a0 &= a(0+0) \\ a0 &= a0 + a0 \\ 0 &= a0. \end{aligned}$$

   Usando a outra lei distributiva, obtemos:

$$\begin{aligned} 0a &= (0+0)a \\ 0a &= 0a + 0a \\ 0 &= 0a. \end{aligned}$$

2. Note que $ab + a(-b) = 0$. Isto porque, conforme vimos no item anterior, podemos escrever:

$$\begin{aligned} ab + a(-b) &= a\left(b + (-b)\right) \\ ab + a(-b) &= a(0) \\ ab + a(-b) &= 0. \end{aligned}$$

   E pelo axioma 5, temos $ab + \left(-(ab)\right) = 0$, assim:

$$\begin{aligned} ab + a(-b) &= 0 \\ a(-b) &= -(ab) \end{aligned}$$

3. $(-a)(-b) = ab$ é um caso particular do item 2. Temos:

$$\begin{aligned} (-a)(-b) &= -\left(a(-b)\right) && \text{pelo item 2} \\ &= -\left(-(ab)\right) && \text{pelo item 2} \\ &= ab \end{aligned}$$

   pois $-(-x) = x$ é uma consequência do fato de que em qualquer grupo $(u^{-1})^{-1} = u$.

4. Suponhamos que R possua um elemento unidade 1. Assim $a + (-1)a = 1a + (-1)a = \left(1 + (-1)\right)a = 0a = 0$, portanto pelo axioma 5 de anel, temos $(-1)a = -a$.

5. Para demonstrar este item, vou tomar $a = -1$ do item anterior. Deste modo, obtemos $(-1)(-1) = -(-1) = 1$.

□

Agora que demonstramos o Lema 4.52 podemos fazer cálculos com os números negativos e o 0 como fazemos usualmente. Por conveniência, $a + (-b)$ será indicado por $a - b$.

Para demonstrar o próximo lema, vamos utilizar um princípio que embora seja muito simples é extremamente valioso em certas ocasiões. Este princípio diz que se temos 11

objetos para guardar em 10 gavetas, então uma gaveta receberá necessariamente dois objetos. Este princípio é conhecido como *Princípio das Gavetas* (ou *Princípio das Casas dos Pombos*) e enunciamos formalmente assim:

Se n objetos são distribuídos em m gavetas e se $n > m$, então algumas gavetas recebem pelo menos dois objetos.

Lema 4.53. *Um anel de integridade finito é um corpo.*

Demonstração. Pela Definição 4.49, um anel de integridade é um anel comutativo tal que $ab = 0$ se, e somente se, pelo menos a ou b é nulo. Já um corpo, é um anel comutativo com elemento unidade no qual todo elemento não nulo possui um inverso multiplicativo no anel.

Considere D um anel de integridade finito. Queremos mostrar que D é corpo. Para isto, devemos demonstrar:

1. que existe o elemento unidade $1 \in D$ tal que $1a = a$ para todo $a \in D$;
2. qualquer que seja, $a \neq 0 \in D$, existe $b \in D$ tal que $ab = 1$.

Sejam $x_1, x_2, \ldots, x_n$ os elementos de D e suponhamos que $a \neq 0 \in D$. Agora consideremos os elementos $x_1a, x_2a, \ldots, x_na$, obviamente eles estão todos em D. Além disso, todos são distintos! Com efeito, suponhamos que $x_ia = x_ja$ para $i \neq j$, então $(x_i - x_j)a = 0$. Como D é um anel de integridade e $a \neq 0$, temos $x_i - x_j = 0$ e consequentemente $x_i = x_j$ o que contradiz a hipótese. Assim $x_1, x_2, \ldots, x_n$ são n elementos distintos de D, que possui exatamente n elementos. Pelo Princípio das Gavetas estes devem esgotar os elementos de D. Agora, vamos mostrar que 1 é elemento unidade de cada elemento $x_i \in D$. Vimos que podemos representar os elementos $x_i \in D$ como $y = x_ia$. Assim, $y1 = (x_ia)1 = x_i(a1) = x_ia = y$. Sabendo que D é comutativo temos $y = 1y = y1$ e portanto 1 é elemento unidade para todo elemento de D. Para resolver a última parte do lema, vamos usar o fato de $1 \in D$, portanto 1 pode ser escrito como um múltiplo de a, ou seja, existe um $b \in D$ tal que $1 = ba$ o que conclui o lema. □

Corolário 4.54. *Se p é um número primo, então $\mathbb{Z}_p$, o anel dos inteiros* $\bmod p$*, é um corpo.*

Demonstração. Pelo Lema 4.53 basta provar que $\mathbb{Z}_p$ é um anel de integridade, pois ele possui apenas um número finito de elementos. Se $a, b \in \mathbb{Z}_p$ e $ab \equiv 0$, então p divide necessariamente o inteiro ab, e assim, p, sendo primo, divide necessariamente

a ou b. E portanto, temos $a \equiv 0 \mod p$ ou $b \equiv 0 \mod p$. Logo $\mathbb{Z}_p$ é um anel de integridade. □

No começo do capítulo colocamos na lista um exemplo de F, o corpo dos inteiros mod p, p um primo. O corolário acima nos prova porque tal conjunto é um corpo. Vamos agora dar um exemplo de um corpo com estas características. Considere o conjunto $F_7 = \{[0], [1], [2], [3], [4], [5], [6]\}$, observe que todos os elementos com exceção do $[0]$ possui um inverso, que é comutativo.

$$
\begin{aligned}
[1]\,[1] &= [1] = [1]\,[1] \\
[2]\,[4] &= [1] = [4]\,[2] \\
[3]\,[5] &= [1] = [5]\,[3] \\
[6]\,[6] &= [1] = [6]\,[6].
\end{aligned}
$$

Note que quando o número não é primo não podemos garantir que F seja um corpo, vejamos um exemplo, $F_6 = \{[0], [1], [2], [3], [4], [5]\}$. Observe que $[2]$ não possui inverso, pois:

$$
\begin{aligned}
[2]\,[1] &= [2] \\
[2]\,[2] &= [4] \\
[2]\,[3] &= [0] \\
[2]\,[4] &= [2] \\
[2]\,[5] &= [4].
\end{aligned}
$$

Além disso, $[2]\,[3] = [0]$ com $[2] \neq [0]$ e $[3] \neq [0]$.

Definição 4.55. A *característica de um anel de integridade* D é definida como sendo o menor inteiro p tal que $pa = 0$ para algum a em D. Diremos que D é *característica* 0 se a relação $ma = 0$, onde $0 \neq a \in D$ e m é um inteiro, vale apenas se $m = 0$. D é dito de *característica finita* se para algum $a \neq 0$ em D e algum inteiro $m \neq 0$, temos $ma = 0$.

4.3.1 *Homomorfismo*

Assim como fizemos em grupos, podemos introduzir o conceito de homomorfismo, ou seja, uma função tal que $\phi(ab) = \phi(a)\phi(b)$. Como um anel possui duas operações, qual poderia ser a extensão mais natural deste tipo de fórmula do que a:

Definição 4.56. Uma função do anel R no anel R' é dita um *homomorfismo* se

1. $\phi(a+b) = \phi(a) + \phi(b)$;
2. $\phi(ab) = \phi(a)\phi(b)$;

para todos $a, b \in R$.

Lema 4.57. *Se ϕ é um homomorfismo de R em R' , então*

1. $\phi(0) = 0$*;*
2. $\phi(-a) = -\phi(a)$ *para todo* $a \in R$.

Como fizemos em grupos, faremos também em anéis. Definiremos o núcleo de um homomorfismo para anéis. Devemos levar em consideração, que anel tem duas operações, adição e multiplicação, e desta maneira é necessário escolher qual das duas operações deve ser escolhida para ser a base da definição. Esta escolha, deve ser tomada levando em conta que anéis arbitrários formam um grupo abeliano com relação à adição, enquanto que a multiplicação é deixada de certo modo, mais "*livre*". Deste modo, definimos

Definição 4.58. Se ϕ é um homomorfismo de R em R', então o núcleo de ϕ, $I(\phi)$, é o conjunto de todos os elementos $a \in R$ tais que $\phi(a) = 0$, o elemento zero de R'.

Lema 4.59. *Se ϕ é um homomorfismo de R em R' , com núcleo $I(\phi)$, então:*

1. $I(\phi)$ *é um subgrupo de* R *com relação à adição.*
2. *Se* $a \in I(\phi)$ *e* $r \in R$*, então* ar *e* ra *estão em* $I(\phi)$.

E também, tal qual fizemos em grupos, também em anéis definimos monomorfismo e o isomorfismo, que possuem a mesma essência descrita para grupos. Desta maneira, chegamos as seguintes definições

Definição 4.60. Um homomorfismo de R em R' é dito um *monomorfismo* se ele é uma função injetiva.

Definição 4.61. Dois anéis são ditos *isomorfos* se existe um monomorfismo sobrejetivo de um sobre o outro. Um monomorfismo sobrejetivo é também denominado um *isomorfismo*.

E, de maneira natural, chegamos a outro lema para anéis que também guarda as mesmas ideias do Lema 4.35 para grupos.

Lema 4.62. *O homomorfismo de R em R' é um monomorfismo se, e somente se, $I(\phi) = 0$.*

4.3.2 *Ideais e Anéis Quocientes*

Utilizando a ideia de homomorfismo e de núcleo para anéis, vamos procurar construir uma estrutura para anéis semelhante ao subgrupo normal. Assim, partindo do Lema 4.59 chegamos a seguinte definição:

Definição 4.63. Um subconjunto não vazio U de R é dito um ideal (bilateral) de R se:

1. U é um subgrupo de R com relação à adição.
2. Para todo $u \in U$ e $r \in R$, ur e ru estão em U.

A condição 2 afirma que U "absorve" a multiplicação pela direita e pela esquerda por elementos arbitrários do anel. Por esta razão U é usualmente denominado um ideal bilateral. Para nosso caso tratarei simplesmente por ideal.

Um ideal do anel dos inteiros é o conjunto dos números pares, note que quando multiplicamos qualquer inteiro por um número par, obtemos um número par. Com efeito, esta ideia pode ser estendida para qualquer conjunto de múltiplos.

Mas examinemos agora a condição 1, ela nos diz que um ideal U de um anel R é um subgrupo com relação à adição. Assim considere R/U o conjunto de todas as classes laterais distintas de U em R que obtemos quando consideramos U como um subgrupo de R com relação à adição (dissemos simplesmente classe lateral e não classe lateral direita ou classe lateral à esquerda, pelo fato de R ser um grupo abeliano com relação à adição). Sabemos que R/U é da forma $a+U$, com $a \in R$ e que $(a+U)+(b+U) = (a+b)+U$, com $a, b \in R$. Mas também temos que garantir que R/U é um anel, assim é natural definirmos $(a+U)(b+U) = (ab+U)$. Entretanto, precisamos assegurar que isto tem significado, ou seja, devemos mostrar que se $a+U = a'+U$ e $b+U = b'+U$ então, com relação a nossa definição de multiplicação, $(a+U)(b+U) = (a'+U)(b'+U)$ é equivalente a $ab+U = a'b'+U$.

Para isto, tomemos $a+U = a'+U$, $a = a'+u_1$, onde $u_1 \in U$, analogamente, $b = b'+u_2$ onde $u_2 \in U$. Deste modo, temos que:

$$ab = (a'+u_1)(b'+u_2) = a'b' + u_1b' + a'u_2 + u_1u_2$$

com U um ideal de R, $u_1b' \in U$, $a'u_2 \in U$ e $u_1, u_2, \in U$.

Assim, podemos escrever:

$$u_1b' + a'u_2 + u_1u_2 = u_3 \in U.$$

Logo

$$ab = a'b' + u_3.$$

Isso nos permite concluir que $ab+U = a'b'+u_3+U = a'b'+U$, e assim a multiplicação está bem definida.

Agora precisamos mostrar que R/U é um anel, para isto, devemos verificar todos os axiomas de anel. É o que faremos. Vamos tomar $(a+U), (b+U), (c+U) \in R/U$, assim temos:

1. $(a+U)+(b+U) = (a+b)+U$, com $(a+b)+U \in R/U$.

2. $(a+U)+(b+U)=(a+b)+U=(b+a)+U=(b+U)+(a+U)$.
3. $((a+U)+(b+U))+(c+U)=((a+b+U))+(c+U)=(a+b+c+U)=(a+U)+((b+c+U))\,(a+U)+((b+U)+(c+U))$.
4. Existe um elemento $e=U$ em R/U tal que $(a+U)+U=a+U$ para cada $(a+U)\in R/U$.
5. Existe um elemento $-a+U$ em R/U tal que $(a+U)+((-a)+U)=U=e$.
6. $(a+U)(b+U)=(ab+U)$, com $(ab+U)\in R/U$.
7. $(a+U)\,((b+U)(c+U))=(a+U)\,((bc+U))=(abc+U)=((ab+U))\,(c+U)=((a+U)(b+U))\,(c+U)$.
8. $(a+U)\,((b+U)+(c+U))=(a+U)\,((b+c)+U)=(a(b+c)+U)=((ab+ac)+U)=(ab+U)+(ac+U)$ e $((b+U)+(c+U))\,(a+U)=((b+c)+U)\,(a+U)=((b+c)a+U)=((ba+ca)+U)=(ba+U)+(ca+U)$.

E ainda mais, se trata de um anel comutativo, pois R é comutativo, assim $(a+U)(b+U)=ab+U=ba+U=(b+U)(a+U)$. (A recíproca disto é falsa.) Se R possui um elemento unidade 1, então R/U possui um elemento unidade $1+U$. Podemos perceber que existe um homomorfismo sobrejetivo ϕ de R em R/U dado por $\phi(a)=a+U$ para todo $a\in R$, cujo núcleo é exatamente U.

Exemplo 4.7. Examinemos um pouco o mais o ideal dos números pares sobre os inteiros. Vamos representar os inteiros pares por $2\mathbb{Z}=\{2x;\,x\in\mathbb{Z}\}$ e o ideal por $(2\mathbb{Z})$. Assim podemos obter o anel quociente $\mathbb{Z}/(2\mathbb{Z})$, onde seus elementos são classes laterais, da forma $x+2\mathbb{Z}$ com $x\in\mathbb{Z}$. Observamos que $2\mathbb{Z}$ é exatamente o núcleo do homomorfismo sobrejetivo ϕ e portanto representam o elemento *"zero"* do conjunto, assim ϕ leva todo elemento $x\in 2\mathbb{Z}$ no elemento *"zero"* de $\mathbb{Z}/(2\mathbb{Z})$ que chamarei de $\bar{0}$. Todos os outros elementos são congruentes. Note que todos os outros elementos de $\mathbb{Z}$ que não estão em $2\mathbb{Z}$ podem ser escritos como $1+2\mathbb{Z}$, que representarei por $\bar{1}$. Assim, $\mathbb{Z}/(2\mathbb{Z})=\{0+2\mathbb{Z},1+2\mathbb{Z}\}=\{\bar{0},\bar{1}\}$.

Podemos notar uma imensa semelhança entre os conjuntos dos Exemplos 4.2 e o 4.7 que acabamos de definir. Os conjuntos $\mathbb{Z}_2$ e $\mathbb{Z}/(2\mathbb{Z})$ são isomorfos, entretanto é comum estabelecer uma relação de igualdade e escrevermos $\mathbb{Z}_2=\mathbb{Z}/(2\mathbb{Z})=\{[0],[1]\}$.

Podemos resumir estas informações no lema à seguir

Lema 4.64. *Se U é um ideal do anel R, então R/U é um anel e é uma imagem homomorfa de R.*

O teorema a seguir relaciona a ideia de anéis quocientes com homomorfismo, bem semelhante ao que fizemos em grupos.

Teorema 4.65. *Sejam R e R' anéis e ϕ um homomorfismo sobrejetivo de R em R' com núcleo U. Então, R' é isomorfo a R/U. Além do mais, existe uma correspondência biunívoca entre o conjunto dos ideais de R' e o conjunto dos ideais de R que contêm U. Esta correspondência pode ser conseguida associando com um ideal W' de R' o ideal W de R definido por $W = \{x \in R; \phi(x) \in W'\}$. Com W assim definido R/W é isomorfo a R'/W'.*

Vamos agora relacionar ideais e anéis quocientes com corpos, que é um tipo de anel muito importante.

Lema 4.66. *Seja R um anel comutativo com elemento unidade cujos únicos ideais são* (0) *e o próprio R. Então, R é um corpo.*

Definição 4.67. Um ideal $M \neq R$ num anel R é dito um *ideal maximal* de R se sempre que U for um ideal de R tal que $M \subset U \subset R$, então $R = U$ ou $M = U$.

Dizemos que um ideal é um ideal maximal, se é impossível colocar um ideal entre ele e o anel todo. Não há garantias que dado um anel R, ele possua ideais maximais, nem tão pouco que dado um anel R exista somente um ideal maximal.

Vejamos agora um teorema que relaciona ideal maximal e corpo.

Teorema 4.68. *Se R é um anel comutativo com elemento unidade e M é um ideal de R, então M é um ideal maximal de R se, e somente se, R/M é um corpo.*

4.3.3 *Anéis Euclidianos*

Esta subseção serve de introdução à subseção seguinte.

Definição 4.69. Um anel de integridade R é dito um *anel euclidiano* se para todo $a \neq 0$ em R está definido um inteiro não negativo $d(a)$ tal que:

1. Para todos $a, b \in R$, ambos não nulos, $d(a) \leq d(ab)$.

2. Para todos $a, b \in R$, ambos não nulos, existem $t, r \in R$ tais que $a = tb + r$, onde $r = 0$ ou $d(r) < d(b)$.

Os inteiros é um exemplo de anel euclidiano, onde $d(a) = a$ com $a \in \mathbb{Z}$. Agora abordaremos uma relação entre um anel de integridade e um ideal A de um anel R.

Teorema 4.70. *Seja R um anel euclidiano e seja A um ideal de R. Então existe um elemento $a_0 \in A$ tal que A consista exatamente de todos os $a_0 x$ quando x percorre R.*

Demonstração. Se A consiste apenas do elemento 0, basta fazer $a_0 = 0$ e o teorema é válido.

Assim podemos admitir que $A \neq (0)$ e portanto existe um $a \neq 0$ em A. Tomemos um $a_0 \in A$ tal que $d(a_0)$ seja mínimo. (Como d assume valores inteiros não negativos isto é sempre possível.)

Suponhamos que $a \in A$. Pela Definição 4.69 existem $t, r \in R$ tais que $a = ta_0 + r$ onde $r = 0$ ou $d(r) < d(a_0)$. Como $a_0 \in A$ e A é um ideal de R, ta_0 está em A. Combinado com $a \in A$ isto resulta em $a - ta_0 \in A$; mas $r = a - ta_0$, donde $r \in A$. Se $r \neq 0$, então $d(r) < d(a_0)$, dando-nos um elemento r de A para o qual d assume um valor menor que para a_0, em contradição com nossa escolha de a_0 como elemento de A para o qual d assume valor mínimo. Consequentemente, $r = 0$ e $a = ta_0$, o que demonstra o teorema. □

Usaremos a notação $(a) = \{xa; x \in R\}$ para representar o ideal de todos os múltiplos de a.

Definição 4.71. Um anel de integridade R com elemento unidade é um *anel principal* se todo ideal A em R é da forma $A = (a)$ para algum $a \in R$.

Corolário 4.72. *Um anel euclidiano possui um elemento unidade.*

Demonstração. Seja R um anel euclidiano, então R é certamente um ideal de R, de modo que pelo Teorema 4.70 podemos concluir que $R = (u_0)$ para algum $u_0 \in R$. Assim todo elemento em R é um múltiplo de u_0. Portanto, em particular, $u_0 = u_0 c$ para

algum $c \in R$. Se R, então $a = xu_0$ para algum $x \in R$, donde $ac = (xu_0)c = x(u_0c) = a$. E portanto c é o elemento unidade. □

Definição 4.73. Se $a \neq 0$ e b estão num anel comutativo R, então diz-se que a divide b se existe um $c \in R$ tal que $b = ac$. Usaremos o símbolo $a \mid b$ para representar o fato de que a divide b e $a \nmid b$ para significar que a não divide b.

Da Definição 4.73 obtemos os seguintes resultados:

1. Se $a \mid b$ e $b \mid c$, então $a \mid c$.
2. Se $a \mid b$ e $a \mid c$, então $a \mid (b \pm c)$.
3. Se $a \mid b$, então e $a \mid bx$ para todo $x \in R$.

Vejamos a validade das afirmações acima.

1. Como $a \mid b$ então podemos escrever:

$$b = am. \tag{4.4}$$

   E por $b \mid c$ temos:

$$c = bn. \tag{4.5}$$

   Substituindo b da equação (4.4) na equação (4.5), temos:

$$\begin{aligned} c &= bn \\ c &= amn. \end{aligned} \tag{4.6}$$

   Observando a equação (4.6) concluímos que $a \mid c$.

2. Como $a \mid b$ então podemos escrever:

$$b = am. \tag{4.7}$$

   E por $a \mid c$ temos:

$$c = an. \tag{4.8}$$

Somando (subtraindo) as equações (4.7) e (4.8), obtemos:

$$b \pm c = am \pm an$$
$$b \pm c = a(m \pm n) \tag{4.9}$$

E assim, conforme a equação (4.9) $a \mid (b \pm c)$.

3. Este item a verificação é imediata.

Definição 4.74. Se $a, b \in R$, então $d \in R$ é dito um *máximo divisor comum* de a e b se:

1. $d \mid a$ e $d \mid b$.
2. Sempre que $c \mid a$ e $c \mid b$, então $c \mid d$.

Para indicar que d é o máximo divisor comum de a e b, usaremos a notação $d = (a, b)$.

Lema 4.75. *Seja R um anel euclidiano. Então dois elementos quaisquer a e b em R possuem um máximo divisor comum d. Além disso, $d = \lambda a + \mu b$ para certos $\lambda, \mu \in R$.*

Demonstração. Considere $A = \{ra + sb; r, s \in R\}$. Afirmamos que A é um ideal de R. De fato, suponhamos que $x, y \in A$, portanto $x = r_1a + s_1b$, $y = r_2a + s_2b$ e então $x \pm y = (r_1 \pm r_2)a + (s_1 \pm s_2)b \in A$. Analogamente, para todo $u \in R$, $ux = u(r_1a + s_1b) = (ur_1)a + (us_1)b \in A$.

Como A é um ideal de R, pelo Teorema 4.70, existe um elemento $d \in A$ tal que todo elemento em A é um múltiplo de d. Devido ao fato de que $d \in A$ e que todo elemento em A é da forma $ra + sb$, $d = \lambda a + \mu b$ para certos $\lambda, \mu \in R$. Ora, pelo Corolário 4.72, R possui um elemento unidade 1, assim $a = 1a + 0b \in A$, $b = 0a + 1b \in A$. Estando em A são ambos múltiplos de d, donde $d \mid a$ e $d \mid b$.

Suponhamos, por fim, que $c \mid a$ e $c \mid b$, então $c \mid \lambda a$ e $c \mid \mu b$ de modo que c certamente divide $\lambda a + \mu b = d$. Portanto, d satisfaz a todas as condições exigidas para um máximo divisor comum e o lema está demonstrado. □

Definição 4.76. Seja R um anel comutativo com elemento unidade. Um elemento $a \in R$ é uma *unidade* em R se existe um elemento $b \in R$ tal que $ab = 1$.

Observação 4.77. *Não confundir unidade com elemento unidade*! Uma unidade é um elemento cujo inverso também está no anel.

Lema 4.78. *Seja R um anel de integridade com elemento unidade e suponhamos que para $a, b \in R$, $a \mid b$ e $b \mid a$ sejam verdadeiros. Então, $a = ub$ onde u é uma unidade em R.*

Demonstração. Como $a \mid b$, $b = xa$ para algum $x \in R$. Por outro lado, como $b \mid a$, $a = yb$ para algum $y \in R$. Assim $b = x(yb) = (xy)b$, mas estes são elementos de um anel de integridade, de modo que podemos cancelar o b e obter $xy = 1$. Assim y é uma unidade em R e $a = yb$ o que conclui a demonstração. □

Definição 4.79. Seja R um anel comutativo com elemento unidade. Dois elementos a e b em R são ditos *associados* se $b = ua$ para alguma unidade u em R.

Podemos observar que num anel euclidiano dois quaisquer máximos divisores comuns de dois elementos dados são associados.

Definição 4.80. No anel euclidiano R uma não unidade π é dita um *elemento primo* de R se sempre que $\pi = ab$, onde a e b estão em R, então a ou b é uma unidade em R.

Em outras palavras, o elemento primo é um elemento de R que não pode ser fatorado em R de maneira não trivial.

Lema 4.81. *Seja R um anel euclidiano. Então todo elemento em R é uma unidade em R ou pode ser escrito como o produto de um número finito de elementos primos de R.*

O teorema está demonstrado em [6].

Definição 4.82. No anel euclidiano R, a e b em R são ditos *primos entre si (ou relativamente primos)* se seu máximo divisor comum é uma unidade de R.

Temos que todo associado de um máximo divisor comum é um máximo divisor comum, e por outro lado, como 1 é um associado de qualquer unidade, se a e b são primos entre si podemos admitir que $(a, b) = 1$.

Lema 4.83. *Seja R um anel euclidiano. Suponhamos que para $a, b, c \in R$, $a \mid bc$ mas $(a, b) = 1$. Então $a \mid c$.*

Demonstração. Pelo Lema 4.75, o máximo divisor comum de a e b pod ser escrito na forma $\lambda a + \mu b$. Assim pelas nossas suposições $\lambda a + \mu b = 1$. Multiplicando esta igualdade ambos os lados por c, obtemos $\lambda ac + \mu bc = c$. Ora, $a \mid \lambda ac$ e $a \mid \mu bc$, pois por hipótese $a \mid bc$. Logo $a \mid (\lambda ac + \mu bc) = c$. □

Lema 4.84. *Se π é um elemento primo no anel euclidiano R e $\pi \mid ab$, onde $a, b \in R$, então π divide pelo menos a ou b.*

Demonstração. Suponhamos que π não divida a, então $(\pi, a) = 1$. Aplicando o Lema 4.83 concluímos que $\pi \mid b$. □

Teorema 4.85. *Teorema da Unicidade da Fatoração*
Seja R um anel euclidiano e $a \neq 0$ uma não unidade em R. Suponhamos que

$$a = \pi_1 \pi_2 \dots \pi_n = \pi'_1 \pi'_2 \dots \pi'_m$$

onde os π_i e π'_j são elementos primos de R. Então, $n = m$ e cada $\pi_i, 1 \leq i \leq n$ é associado de algum π'_j, $1 \leq j \leq m$ e reciprocamente, cada π'_k é um associado de algum π_q.

Demonstração. Observando que $a = \pi_1 \pi_2 \dots \pi_n = \pi'_1 \pi'_2 \dots \pi'_m$, $\pi_1 \mid \pi_1 \pi_2 \dots \pi_n$, portanto $\pi_1 \mid \pi'_1 \pi'_2 \dots \pi'_m$. Pelo Lema 4.84, π_1 divide algum π'_i. Por outro lado, π_1 e π'_i são ambos elementos primos de R e $\pi \mid \pi'_i$ eles são associados e $\pi'_i = u_1 \pi_1$ onde u_1 é uma unidade de R. Deste modo, $\pi_1 \pi_2 \dots \pi_n = \pi'_1 \pi'_2 \dots \pi'_m = u_1 \pi_1 \pi'_2 \dots \pi'_{i-1} \pi'_{i+1} \dots \pi'_m$ e cancelando π_1 ficamos com $\pi_2 \dots \pi_n = u_1 \pi'_2 \dots \pi'_{i-1} \pi'_{i+1} \dots \pi'_m$. Repetimos o mesmo argumento com π_2. Após alguns passos, o primeiro membro torna-se 1, o segundo membro, o produto de um certo número de π' (o excesso de m sobre n). Isto implicaria $n \leq m$ pois os π' não são unidades. Analogamente, $m \leq n$, de modo que $n = m$. Assim, concluímos que todo π'_i possui algum π'_j como um associado e reciprocamente, todo π'_j possui algum π'_i como um associado. □

Associando o Lema 4.81 e o Teorema 4.85 temos que todo elemento não nulo num anel euclidiano R pode ser escrito de maneira única (a menos de associados) como um produto de elementos primos ou é uma unidade em R.

Investiguemos agora, uma maneira de obter ideais maximais em um anel euclidiano.

Lema 4.86. *O ideal $A = (a_0)$ é um ideal maximal do anel euclidiano R se, e somente se, a_0 é um elemento primo de R.*

Demonstração. Primeiro vamos mostrar que se a_0 não for um elemento primo, então $A = (a_0)$ não é um ideal maximal. De fato, suponhamos que $a_0 = bc$, onde $b, c \in R$ e nem b e nem c são unidades. Considere $B = (b)$, então certamente $a_0 \in B$ de modo que $A \subset B$. Afirmamos que $A \neq B$ e que $B \neq R$.

Se $B = R$, então $1 \in B$ de modo que $1 = xb$ para algum $x \in R$, implicando que b é uma unidade em R, o que é absurdo. Por outro lado, se $A = B$, então $b \in B = A$ donde $b = xa_0$ para algum $x \in R$. Isto, juntamento com o fato de que $a_0 = bc$ resulta em $a_0 = xca_0$, e consequentemente $xc = 1$. Mas isto implica que c é uma unidade em R, contradizendo novamente nossa hipótese. Portanto, B é diferente de A e R e, como $A \subset B$, A não pode ser um ideal maximal de R.

Reciprocamente, suponhamos que a_0 seja um elemento primo de R e que U seja um ideal de R tal que $A = (a_0) \subset U \subset R$. Pelo Teorema 4.70, $U = (u_0)$. Como $a_0 \in A \subset U = (u_0)$, $a_0 = xu_0$ para algum $x \in R$. Mas a_0 é um elemento primo de R, assim temos x ou u_0 uma unidade em R. Pela nossa hipótese, U é um ideal de R que contém uma unidade, portanto pelo Lema 4.81 todo elemento de R é uma unidade em R e portanto $U = R$. Se, por outro lado, x é uma unidade em R, então $x^{-1}R$ e a relação $a_0 = xu_0$ torna-se $x^{-1}a_0 \in A$ pois A é um ideal de R. Isto implica que $U \subset A$, junto com $A \subset U$ concluímos que $U = A$. Portanto, não existe nenhum ideal de R que esteja estritamente entre A e R. Isto significa que A é um ideal maximal de R. □

4.3.4 *Anéis de Polinômios*

Seja F um corpo, definimos *anel de polinômios* na indeterminada x, indicado por $F[x]$, o conjunto de todos os símbolos $a_0 + a_l x + \cdots + a_n x^n$, onde n pode ser qualquer inteiro não negativo e onde os coeficientes $a_0, a_1, \ldots, a_n$ estão todos em F. Para tornar $F[x]$ um anel precisamos reconhecer quando dois elementos dele são iguais, precisamos adicionar e multiplicar elementos de $F[x]$ de modo que os axiomas que definem um anel valham para $F[x]$. É o que faremos.

Definição 4.87. Se $p(x) = a_0 + a_1 x + \cdots + a_m x^m$ e $q = b_0 + b_1 x + \cdots + b_n x^n$ estão em $F[x]$, então $p(x) = q(x)$ se, e somente se, para todo inteiro $i \geq 0, a_i = b_i$.

Assim dois polinômios são ditos iguais se, e somente se, seus coeficientes correspondentes são iguais.

Definição 4.88. Se $p(x) = a_0 + a_1x + \cdots + a_mx^m$ e $q = b_0 + b_1x + \cdots + b_nx^n$ estão ambos em $F[x]$, então $p(x) + q(x) = c_0 + c_1x + \cdots + c_tx^t$, onde para cada i, $c_i = a_i + b_i$.

Definição 4.89. Se $p(x) = a_0 + a_1x + \cdots + a_mx^m$ e $q = b_0 + b_1x + \cdots + b_nx^n$, então $p(x)q(x) = c_0 + c_1x + \cdots + c_kx^k$, onde $c_t = a_tb_0 + a_{t-1}b_1 + a_{t-2}b_2 + \cdots + a_0b_t$.

Apesar de não aparecer a primeira vista, esta multiplicação é a que fazemos usualmente, ou seja, usamos a relação $x^\alpha x^\beta = x^{\alpha+\beta}$ e operamos os termos semelhantes.

Agora podemos afirmar formalmente que $F[x]$ é um anel com as operações de adição (Definição 4.88) e multiplicação (Definição 4.89), pois dados $f(x), g(x), h(x) \in F[x]$, onde $f(x) = \sum_{i=0}^{m} a_ix^i$, $g(x) = \sum_{j=0}^{n} b_jx^j$ e $h(x) = \sum_{k=0}^{r} c_kx^k$, temos:

1. $f(x) + g(x) \in F[x]$. Pois, supondo sem perda de generalidade que $m > n$, podemos escrever $m = n + p$. Assim reescrevemos $g(x) = b_0 + b_1x + b_2x^2 + \cdots + b_nx^n + b_{n+1}x^{n+1} + \cdots + b_mx^m$, onde $b_{n+1} = \cdots = b_m = 0$. E desta forma, obtemos $f(x) + g(x) = (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_m + b_m)x^m \in F[x]$.

2. $f(x) + g(x) = g(x) + f(x)$. No item anterior encontramos, $f(x) + g(x) = (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_m + b_m)x^m \in F[x]$. De modo análogo, chegamos em $g(x) + f(x) = (b_0 + a_0) + (b_1 + a_1)x + \cdots + (b_m + a_m)x^m \in F[x]$, como $a_i, b_j \in F$, temos $(a_0 + b_0) = (b_0 + a_0)$, $(a_1 + b_1) = (b_1 + a_1)$, ..., $(a_m + b_m) = (b_m + a_m)$. Logo $f(x) + g(x) = g(x) + f(x)$.

3. $(f(x) + g(x)) + h(x) = f(x) + (g(x) + h(x))$. Usando os argumentos dos itens anteriores, podemos escrever $f(x) + g(x) = (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_m + b_m)x^m$. Analogamente obtemos $(f(x) + g(x)) + h(x) = [(a_0 + b_0) + c_0] + [(a_1 + b_1) + c_1]\,x + \cdots + [(a_m + b_m) + c_m]\,x^m$. Por outro lado, temos $[(a_0 + b_0) + c_0] + [(a_1 + b_1) + c_1]\,x + \cdots + [(a_m + b_m) + c_m]\,x^m = [a_0 + (b_0 + c_0)] + [a_1 + (b_1 + c_1)]\,x + \cdots + [a_m + (b_m + c_m)]\,x^m = f(x) + (g(x) + h(x))$.

4. Chamamos de *polinômio nulo* o polinômio $O(x) = 0$. Deste modo, $f(x) + O(x) = f(x)$ para todo $f(x) \in F[x]$.

5. Existe $-f(x) \in F[x]$ tal que $f(x) + (-f(x)) = O(x)$. Pois, $-f(x) = -(a_0 + a_1x + a_2x^2 + \cdots + a_mx^m) = -a_0 - a_1x - a_2x^2 - \cdots - a_mx^m$. Assim $f(x) + (-f(x)) = (a_0 - a_0) + (a_1 - a_1)x + \cdots + (a_m - a_m)x^m = 0 = O(x)$.

6. $f(x)g(x) \in F[x]$. Usando a Definição 4.89, obtemos:

$$f(x)g(x) = d_0 + d_1x + \cdots + d_sx^s,$$

onde $d_t = a_tb_0 + a_{t-1}b_1 + a_{t-2}b_2 + \cdots + a_0b_t$. E portanto, claramente $f(x)g(x) \in F[x]$.

7. $(f(x)g(x))\,h(x) = f(x)\,(g(x)h(x))$. Vamos primeiramente efetuar $f(x)g(x)$. Como vimos no item anterior, temos $f(x)g(x) = d_0 + d_1x + \cdots + d_sx^s$, onde $d_t = a_tb_0 + a_{t-1}b_1 + a_{t-2}b_2 + \cdots + a_0b_t$. Pela Definição 4.89, o elemento d_t está relacionado ao termo d_tx^t. Como a multiplicação na variável x é obtida somando os expoentes de x, o coeficiente associado a x^t é obtido somando todas as multiplicações dos coeficientes a_i de $f(x)$ por b_j de $g(x)$, onde $i + j = t$. Assim,

$$f(x)g(x) = d_0 + d_1x + \cdots + d_sx^s = \sum_{\alpha=0}^{s} d_\alpha x^\alpha = \sum_{i=0,j=0}^{m,n} a_ib_jx^{i+j}. \tag{4.10}$$

Agora, tomando $f(x)g(x) = \sum_{\alpha=0}^{s} d_\alpha x^\alpha$ e multiplicando por $h(x) = \sum_{k=0}^{r} c_kx^k$, de modo análogo ao que fizemos anteriormente, encontramos:

$$\left(\sum_{\alpha=0}^{s} d_\alpha x^\alpha\right)\left(\sum_{k=0}^{r} c_kx^k\right) = \sum_{\alpha=0,k=0}^{s,r} d_\alpha c_k x^{\alpha+k}. \tag{4.11}$$

Utilizando as equações (4.10) e(4.11), podemos enfim mostrar a igualdade desejada.

$$
\begin{aligned}
\left(f(x)g(x)\right)h(x) &= \left(\sum_{\alpha=0}^{s} d_\alpha x^\alpha\right)\sum_{k=0}^{r} c_k x^k \\
&= \sum_{\alpha=0,k=0}^{s,r} d_\alpha c_k x^{\alpha+k} \\
&= \sum_{\alpha=0,k=0}^{s,r} d_\alpha x^\alpha c_k x^k \\
&= \sum_{i=0,j=0,k=0}^{m,n,r} a_i b_j x^{i+j} c_k x^k \\
&= \sum_{i=0,j=0,k=0}^{m,n,r} a_i x^i b_j x^j c_k x^k \\
&= \left(\sum_{i=0}^{m} a_i x^i\right)\left(\sum_{j=0,k=0}^{n,r} b_j c_k x^{j+k}\right) \\
&= f(x)\left(g(x)h(x)\right).
\end{aligned}
$$

8. $f(x)\left(g(x)+h(x)\right) = f(x)g(x)+f(x)h(x)$ e $\left(g(x)+h(x)\right)f(x) = g(x)f(x)+h(x)f(x)$. Utilizando sem perda de generalidade o resultado de $f(x)+g(x)$ para $g(x)+h(x)$ apresentado no item 1, $n > r$ e tomando $h(x) = c_0 + c_1x + c_2x^2 + \cdots + c_rx^r + c_{r+1}x^{r+1} + \cdots + c_nx^n$, onde $c_{r+1} = \cdots = c_n = 0$. Para facilitar a notação adotarei o índice u, onde $u = \{0, 1, \ldots, n\}$, com $b_u = b_j$, $c_u = c_k$ e $c_u = c_{r+q}$, com $q = n - r$. Desta forma, obtemos:

$$
\begin{aligned}
f(x)\left(g(x)+h(x)\right) &= f(x)\left((b_0+c_0)+(b_1+c_1)x+\cdots+(b_n+c_n)x^n\right) \\
&= \sum_{i=0}^{m} a_i x^i \left(\sum_{u=0}^{n}(b_u+c_u)x^u\right) \\
&= \sum_{i=0,u=0}^{m,n} \left[a_i(b_u+c_u)\right] x^{i+u} \\
&= \sum_{i=0,u=0}^{m,n} \left(a_ib_u + a_ic_u\right) x^{i+u} \\
&= \sum_{i=0,u=0}^{m,n} a_ib_u x^{i+u} + \sum_{i=0,u=0}^{m,n} a_ic_u x^{i+u} \\
&= f(x)g(x) + f(x)h(x).
\end{aligned}
$$

A outra distributividade é feita de modo análogo.

Definição 4.90. h Se $f(x) = a_0 + a_1x + \cdots + a_nx^n \neq 0$ e $a_n \neq 0$, então o *grau* de $f(x)$, indicado por $gr\ f(x)$, é n.

Isto é, o grau de $f(x)$ é o maior inteiro i para o qual o i-ésimo coeficiente de $f(x)$ não é 0. Não definimos o grau do polinômio nulo. Dizemos que um polinômio é uma constante se for um polinômio diferente de zero com grau igual zero. A função-grau definida para os elementos não nulos de $F[x]$ nos fornecerá a função $d(x)$ necessária para que $F[x]$ seja um anel euclidiano.

Lema 4.91. *Se $f(x)$, $g(x)$ são dois elementos não nulos de $F[x]$, então* gr $(f(x)g(x)) =$ gr $f(x) +$ gr $g(x)$.

Demonstração. Sejam $f(x) = \sum_{i=0}^{m} a_ix^i$ e $g(x) = \sum_{j=0}^{n} b_jx^j$, com i e j não todos nulos. Fazendo a multiplicação e considerando sem perda de generalidade $a_m \neq 0$ e $b_n \neq 0$ obtemos, $f(x)g(x) = \sum_{i=0,u=0}^{m,n} a_ib_jx^{i+j}$. Desta forma, temos gr $f(x)g(x) = m + n =$ gr $f(x) +$ gr $g(x)$. □

Corolário 4.92. *Se $f(x) \neq 0$ e $g(x) \neq 0$ são elementos em $F[x]$, então* gr $f(x) \leq$ gr $f(x)g(x)$.

Demonstração. Como $f(x)$ e $g(x)$ são elementos não nulos em $F[x]$, temos as seguintes possibilidades:

1. Se $f(x)$ e $g(x)$ forem funções constantes, então $m = n = 0$. Assim gr $f(x) = 0 = 0 + 0 =$ gr $f(x) +$ gr $g(x)$.
2. Se $f(x)$ ou $g(x)$ for uma função constante e a outra uma função de grau $k > 0$, temos:
   - Caso $f(x)$ seja a função constante, temos gr $f(x) = 0 \leq 0 + k = k =$ gr $f(x) +$ gr $g(x)$.
   - Caso $g(x)$ seja a função constante, temos gr $f(x) = k = k + 0 = k =$ gr $f(x) +$ gr $g(x)$.
3. Se $f(x)$ e $g(x)$ são funções de graus $m > 0$ e $n > 0$ respectivamente, então temos gr $f(x) = m \leq m + n =$ gr $f(x) +$ gr $g(x)$.

□

Corolário 4.93. *$F[x]$ é um anel de integridade.*

Demonstração. $F[x]$ é um anel comutativo e claramente não possui divisores para o polinômio nulo, além do próprio polinômio nulo. □

O fato de $F[x]$ ser um anel de integridade, nos permite com o auxílio do Teorema 4.85 construir para ele o seu corpo de frações. Este corpo consiste de todas as frações polinomiais e é denominado corpo de *funções racionais* em x sobre F.

A função gr $f(x)$ definida para todo $f(x) \neq 0$ em $F[x]$ satisfaz:

1. gr $f(x)$ é um inteiro não negativo.
2. gr $f(x) \leq$ gr $f(x)g(x)$ para todo $g(x) \neq 0$ em $F[x]$.

Para que $F[x]$ seja um anel euclidiano com a função-grau funcionando como a função d de um anel euclidiano ainda precisamos mostrar que dados $f(x), g(x) \in F[x]$ existem $t(x)$, $r(x)$ em $F[x]$ tais que $f(x) = t(x)g(x) + r(x)$ onde $r(x) = 0$ ou gr $r(x) <$ gr $g(x)$. Faremos isto por intermédio do lema a seguir.

Lema 4.94. *O algoritmo da divisão*
Dados dois polinômios $f(x)$ e $g(x) \neq 0$ em $F[x]$, então existem dois polinômios $t(x)$ e $r(x)$ em $F[x]$ tais que $f(x) = t(x)g(x) + r(x)$ onde $r(x) = 0$ ou gr $r(x) <$ gr $g(x)$.

A demonstração deste lema pode ser encontrado em [6], e como o próprio nome diz se trata do processo das "divisões sucessivas", onde dividimos um polinômio por outro.

Agora, com tudo que foi exposto acima podemos enunciar o teorema seguinte.

Teorema 4.95. *$F[x]$ é um anel euclidiano.*

Usando o fato de $F[x]$ ser um anel euclidiano, os resultados da Subseção 4.3.3 são válidos. Assim, podemos relacionar os Lemas 4.96, 4.98 e 4.99, que serão enunciados a seguir, aos já demonstrados. No entanto, também podemos encontrar as demonstrações dos lemas em [4].

Lema 4.96. *Dados dois polinômios $f(x), g(x)$ em $F[x]$ eles possuem um máximo divisor comum $d(x)$ que pode ser representado como $d(x) = \lambda(x)f(x) + \mu(x)g(x)$.*

Definição 4.97. Um polinômio $p(x)$ em $F[x]$ é dito *irredutível* sobre F se sempre que $p(x) = a(x)b(x)$, com $a(x), b(x) \in F[x]$, então $a(x)$ ou $b(x)$ tem grau 0, ou seja, é uma constante.

Lema 4.98. *Qualquer polinômio em $F[x]$ pode ser escrito de uma única maneira como um produto de polinômios irredutíveis em $F[x]$.*

Lema 4.99. *O ideal $A = (p(x))$ em $F[x]$ é um ideal maximal se, e somente se, $p(x)$ é irredutível sobre $F[x]$.*

5

ESPAÇO VETORIAL

O *espaço vetorial* é uma estrutura algébrica semelhante aos anéis e corpos, pois possui duas operações, mas se diferencia pelo fato de que a multiplicação é estabelecida em um conjunto externo conforme veremos adiante.

5.1 ESPAÇOS VETORIAIS

Antes de definirmos espaço vetorial, é importante dizer que este assunto é de grande importância na matemática além de ter diversas aplicações. Para estudá-lo não é pré requisito o estudo de grupos e anéis, mas é necessário conhecer a definição de corpo. O campo que estuda espaços vetoriais é conhecido como *Álgebra Linear* e o leitor pode encontrar mais a respeito em [5].

Definição 5.1. Um conjunto não vazio V é dito um *espaço vetorial* sobre um corpo F se V é um grupo abeliano com relação a uma operação que indicamos por $+$, e se para todos $\alpha, \beta \in F$ e $v, w \in V$ está definido um elemento, indicado por αv, em V, tal que:

1. $\alpha(v + w) = \alpha v + \alpha w$;
2. $(\alpha + \beta)v = \alpha v + \beta v$;
3. $\alpha\,(\beta v) = (\alpha\beta)\,v$;
4. $1v = v$;

para todos $\alpha, \beta \in F$ e $v, w \in V$ (onde 1 representa o elemento unidade de F com relação à multiplicação).

Na definição é dito que V é um espaço vetorial sobre F se V é um grupo abeliano com relação a uma operação que indicamos por $+$, ou seja, temos que:

1. Para todos os elementos u, v em V implica que $u + v = v + u$ (comutativo).
2. Para todos os elementos $u, v, w \in V$ implica que $u + (v + w) = (u + v) + w$ (lei associativa).
3. Existe um elemento $0 \in V$ tal que $v + 0 = v$ para todo $v \in V$ (existência de um elemento unidade em V que no caso é o elemento neutro).
4. Para todo $v \in V$ existe um elemento $-v \in V$ tal que $v + (-v) = 0$ (existência de inversos em V, neste caso, inversos aditivos que também são chamados de simétricos).

Os elementos de V serão chamados de *vetores* e os elementos de F de *escalares*. O elemento 0 de V será representado simplesmente por 0, para facilitar a notação, e é chamado de *vetor nulo* . Denominamos de *vetor oposto* de v o vetor $-v$.

Podemos citar como exemplos importantes e de certo modo de fácil visualização os espaços de $\mathbb{R}$ e $\mathbb{C}$ sobre o corpo $\mathbb{Q}$ e o espaço $\mathbb{C}$ sobre o corpo $\mathbb{R}$.

O próximo lema nos dará mais informações sobre um espaço vetorial, sua demonstração pode ser encontrada em [6].

Lema 5.2. *Se V é um espaço vetorial sobre F, então:*

1. $\alpha 0 = 0$ *para* $\alpha \in F$.
2. $0v = 0$ *para* $v \in V$.
3. $(-\alpha)v = -(\alpha v)$ *para* $\alpha \in F, v \in V$.
4. $v \neq 0$, *então* $\alpha v = 0$ *implica que* $\alpha = 0$.

Vejamos alguns exemplos de especial interesse para nós.

Exemplo 5.1. Sejam F um corpo e V a totalidade das ênuplas ordenadas $(\alpha_1, \ldots, \alpha_n)$, onde $\alpha_i \in F$. Dois elementos $(\alpha_1, \ldots, \alpha_n)$ e $(\beta_1, \ldots, \beta_n)$ são definidos como iguais se, e somente se, $\alpha_i = \beta_i$ para cada $i \in \{1, \ldots, n\}$. Definimos também:

1. $(\alpha_1, \ldots, \alpha_n) + (\beta_1, \ldots, \beta_n) = (\alpha_1 + \beta_1, \alpha_2 + \beta_2, \ldots, \alpha_n + \beta_n)$.
2. $\gamma\,(\alpha_1, \ldots, \alpha_n) = (\gamma\alpha_1, \ldots, \gamma\alpha_n)$ para $\gamma \in F$.

Da maneira como as operações em V foram definidas é fácil verificar que V é um espaço vetorial sobre F. Denotaremos o espaço vetorial que acabamos de definir por $F^{(n)}$.

Exemplo 5.2. Consideremos F um corpo qualquer e $V = F[x]$, o conjunto dos polinômios em x sobre F. Tomando a adição usual entre polinômios e a multiplicação usual de um elemento α em F por algum $p(x)$ em $F[x]$, verificamos que $F[x]$ é um espaço vetorial sobre F.

Exemplo 5.3. Seja $V_n \subset F[x]$ o conjunto de todos os polinômios de graus menores do que n. Tomando as operações usuais de adição e multiplicação para polinômios, V_n é um espaço vetorial sobre F.

5.2 SUBESPAÇO VETORIAL

Definição 5.3. Se V é um espaço vetorial sobre F e se $W \subset V$, então W é um *subespaço vetorial* de V se, com relação às operações de V, W forma um espaço vetorial sobre F.

Para determinar se um subconjunto não vazio de um espaço vetorial é um espaço vetorial não será necessário verificar todos os axiomas de espaço vetorial. Considere $W \subset V$ e V espaço vetorial. Então pela Definição 5.3, W é *subespaço vetorial* de V ou simplesmente *subespaço* de V, se com relação às operações de V, W forma um espaço vetorial. Contudo, como W é subconjunto de V e V é espaço vetorial, sabemos que as operações são comutativas e associativas em todos os elementos de V, logo todos os elementos de W também gozam destas propriedades, ou seja, para quaisquer $w_1, w_2, w_3 \in W$ temos $w_1 + w_2 = w_2 + w_1$ e $(w_1 + w_2) + w_3 = w_1 + (w_2 + w_3)$.

Também podemos concluir que as duas distributividades, a associatividade com relação à $\alpha, \beta \in F$ e $w \in W$ e a igualdade $1w = w$ são naturalmente cumpridas pois $W \subset V$.

Para assegurar que um subconjunto não vazio W de V, onde V é um espaço vetorial, seja um subespaço de V, nos falta mostrar que existe elemento neutro em W e que cada $w \in W$ possui simétrico. O Teorema 5.4 a seguir resume estas informações.

Teorema 5.4. *Sejam V um espaço vetorial e W um subconjunto não vazio de V . Temos que W é um subespaço vetorial de V se, e somente se, $w_1 + \alpha w_2 \in W$, para todo $\alpha \in F$ e para todos $w_1, w_2 \in W$.*

Demonstração. Suponhamos que W seja subespaço de V. Assim dados $w_1, w_2 \in W$ e $\alpha \in F$, temos que $\alpha w_2 \in W$. Agora, como $w_1, \alpha w_2 \in W$ então $w_1 + \alpha w_2 \in W$. Reciprocamente, supondo que para quaisquer $w_1, w_2 \in W$ e $\alpha \in F$, temos $w_1 + \alpha w_2 \in W$. Devemos mostrar que W é um espaço vetorial. Para isto, basta provar que existe elemento neutro em W e que cada $w \in W$ possui simétrico. Assim, tomemos $\alpha = -1$. Desta forma,

$$w + \alpha w = w + (-1)w = w + (-w) = 0 \in W.$$

Logo $0 \in W$ e $w \in W$. Agora, para $\alpha = -1$, temos $0 + (-1)w = 0 + (-w) = -w \in W$ e portanto existe um elemento neutro em W e cada vetor w em W possui um vetor oposto $-w$ em W.

□

Um exemplo importante de subespaço vetorial é o espaço vetorial formado somente pelo vetor nulo. Este subespaço é chamado *subespaço vetorial nulo*. O próprio espaço vetorial V, também é subespaço de V.

5.2.1 *Operações com Subespaços*

Os subespaços vetoriais são conjuntos e desta maneira podemos nos perguntar: A união de dois subespaços vetoriais, é um espaço vetorial? E a interseção? Vamos responder estas questões.

Para responder a primeira pergunta, podemos usar um contra-exemplo. Vamos fazer a união de dois subespaços e mostrar que não resulta em um conjunto que preserva as propriedades de espaço vetorial.

Exemplo 5.4. Dados $U = \{(x, y) \in \mathbb{R}^2; x + y = 0\}$ e $W = \{(x, y) \in \mathbb{R}^2; x - y = 0\}$, subespaços de $\mathbb{R}^2$, o conjunto $U \cup W$ não é um subespaço de $\mathbb{R}^2$. Com efeito, temos que $u = (1, 1) \in U \cup W$ e $w = (1, -1) \in U \cup W$, mas $u + w = (2, 0) \notin U \cup W$.

Com isto, mostramos que não há garantias de que a união de dois subespaços de um espaço vetorial V seja um subespaço de V.

Agora nos voltemos a segunda questão. E quanto a interseção? Neste caso, sim. Podemos provar que a interseção de dois subespaços vetoriais de V é necessariamente um subespaço de V. Isto nos leva à seguinte proposição:

Proposição 5.5. *A interseção de dois subespaços de um espaço vetorial V é um subespaço de V.*

Demonstração. Sejam U e W subespaços de V. Para verificarmos que $U \cap W$ é também um subespaço de V, vamos fazer uso do Teorema 5.4. Para isto, primeiramente note que $U \cap W$ é um subconjunto não vazio de V , pois $0 \in U$ e $0 \in W$, já que ambos U e W são subespaços de V. Agora, tomemos $\alpha \in F$ e $u, w \in U \cap W$. Como $u, w \in U$ e $u, w \in W$, segue do Teorema 5.4 que $u + \alpha w \in U$ e $u + \alpha w \in W$, ou seja, $u + \alpha w \in U \cap W$. Novamente, pelo Teorema 5.4, segue que $U \cap W$ é um subespaço de V. □

Vamos definir agora a soma de dois subespaços vetoriais de V.

Definição 5.6. Sejam U e W subespaços de um espaço vetorial V , definimos a soma de U e W, denotada por $U + W$, como o conjunto:

$$U + W = \{u + w;\ u \in U \text{ e } w \in W\}.$$

Agora, vamos mostrar que a soma de dois subespaços de V é um subespaço de V.

Proposição 5.7. *A soma de dois subespaços U e W de um espaço vetorial V é um subespaço de V.*

Demonstração. Sejam U e W subespaços de V. Tomemos $\alpha \in F$ e $v_1, v_2 \in U + W$. Como $v_1, v_2 \in U + W$, existem $u_1, u_2 \in U$ e $w_1, w_2 \in W$ tais que $v_1 = u_1 + w_1$ e $v_2 = u_2 + w_2$. Deste modo, podemos escrever:

$$v_1 + \alpha v_2 = (u_1 + w_1) + \alpha(u_2 + w_2) = (u_1 + \alpha u_2) + (w_1 + \alpha w_2) \in U + W,$$

pois $u_1 + \alpha u_2 \in U$ e $w_1 + \alpha w_2 \in W$. Logo $U + W$ é um subespaço de V. □

Vejamos um exemplo de soma de dois subespaços vetoriais do $\mathbb{R}^2$.

Exemplo 5.5. Considere os subespaços $U = \{(x, y) \in \mathbb{R}^2; x + y = 0\}$ e $W = \{(x, y) \in \mathbb{R}^2; x - y = 0\}$ de $\mathbb{R}^2$. Vamos efetuar $U + V$. Temos:

$$\begin{aligned} (x, y) \in U &\iff x + y = 0 \\ &\iff x = -y. \end{aligned}$$

Deste modo $(-y, y) \in U$, $y \in \mathbb{R}$. Agora tomemos $\alpha, \beta \in \mathbb{R}$, tal que $y = \frac{\alpha-\beta}{2}$. Assim temos $-y = \frac{-\alpha+\beta}{2}$, Logo:

$$\left(\frac{-\alpha+\beta}{2}, \frac{\alpha-\beta}{2}\right) \in U.$$

Analogamente, temos:

$$\begin{aligned} (x, y) \in W &\iff x - y = 0 \\ &\iff x = y. \end{aligned}$$

Deste modo, $(x, x) \in W$, $x \in \mathbb{R}$. Tomando $x = \frac{\alpha+\beta}{2}$, temos:

$$\left(\frac{\alpha+\beta}{2}, \frac{\alpha+\beta}{2}\right) \in W.$$

Assim

$$\left(\frac{-\alpha+\beta}{2}, \frac{\alpha-\beta}{2}\right) + \left(\frac{\alpha+\beta}{2}, \frac{\alpha+\beta}{2}\right) = (\beta, \alpha).$$

Portanto podemos escrever qualquer elemento $(\beta, \alpha) \in \mathbb{R}^2$ como sendo a soma de um elemento $u \in U$ com um elemento $w \in W$. Logo $U + W = \mathbb{R}^2$.

Vamos definir agora um caso particular de soma de subespaços, a *soma direta*.

Definição 5.8. Dados U e W subespaços de um espaço vetorial V. Dizemos que o espaço vetorial V é a *soma direta* de U e W, e representamos por $V = U \oplus W$, se $V = U + W$ e $U \cap W = \{0\}$.

Teorema 5.9. *Sejam U e W subespaços de um espaço vetorial V sobre F. Temos que $V = U \oplus W$ se, e somente se, todo vetor v em V se escreve de modo único como $v = u + w$, onde $u \in U$ e $w \in W$.*

Demonstração. Suponhamos $V = U \oplus W$. Tomemos $v \in V$, pela definição de soma de subespaços podemos escrever $v = u + w$, onde $u \in U$ e $w \in W$. E podemos dizer ainda mais, pela definição esta escrita é única, ou seja, se escrevermos $v = u' + w'$ com $u' \in U$ e $w' \in W$ então $u = u'$ e $w = w'$. Com efeito, se $v = u + w$ e $v = u' + w'$, então:

$$u + w = u' + w' \Longrightarrow \underbrace{u - u'}_{\in U} = \underbrace{w' - w}_{\in W}.$$

Logo $u - u' \in U \cap W$. Como $U \cap W = \{0\}$, acarreta que $u - u' = 0 \iff u = u'$. Analogamente $w' - w \in U \cap W = \{0\}$. O que implica em $w' - w = 0 \iff w' = w$. Reciprocamente, vamos supor que todo vetor de V seja escrito de forma única como a soma de um vetor de U por um vetor de W. Deste modo, $V = U + W$. Agora suponhamos por absurdo que $U \cap W \neq \{0\}$. Deste modo, existe um vetor v não nulo em $U \cap W$. Tal vetor $v \in W$, e por W ser subespaço, temos que $-v \in W$ também. A consequência disto é que, desta maneira podemos escrever $0 = 0 + 0$ com $0 \in U$ e $0 \in W$. Mas também podemos escrever $0 = v + (-v)$, com $v \in U$ e $-v \in W$. Como $v \neq 0$, temos duas escritas distintas para um mesmo vetor v o que contradiz a nossa hipótese. Logo $U \cap W = \{0\}$. □

Podemos apresentar como exemplo de soma direta, a soma dos subespaços U e W do $\mathbb{R}^2$ do Exemplo 5.5. Temos que os elementos de U são da forma $(x, -x)$ e de W da forma (x, x), assim o único elemento que pertence aos dois conjuntos é o vetor nulo, logo $U \cap W = \{0\}$. Portanto $\mathbb{R}^2 = U \oplus W$.

5.3 ESPAÇOS GERADOS

Uma maneira de se obter um espaço vetorial sobre F é através de um conjunto $\{v_1, v_2, \ldots, v_r\}$ de vetores de um espaço vetorial V. Para isto, vamos tomar cada elemento do conjunto W que queremos *gerar* como sendo uma *combinação linear* de vetores $v_1, v_2, \ldots, v_r$ do espaço vetorial V. Isto é:

Definição 5.10. Dado um vetor w em W se existirem $\alpha_1, \alpha_2, \ldots, \alpha_r$ em F, onde F é um corpo, tal que $w = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_r v_r$, então diremos que w em W é uma *combinação linear* sobre F de $v_1, v_2, \ldots, v_r$.

Por simplicidade, por estarmos tratando com um corpo fixado F, frequentemente diremos apenas combinação linear, em vez de combinação linear sobre F. Da mesma forma usaremos espaço vetorial, em vez de espaço vetorial sobre F.

Afirmamos que o conjunto W formado por todas as combinações lineares dos vetores $v_1, v_2, \ldots, v_r$ é *subespaço gerado* por $v_1, v_2, \ldots, v_r$ e dizemos que $v_1, v_2, \ldots, v_r$ *geram* W ou que $\{v_1, v_2, \ldots, v_r\}$ é um *conjunto gerador* de W. Para indicarmos que W é o espaço gerado por $v_1, v_2, \ldots, v_r$, escrevemos $W = G(v_1, v_2, \ldots, v_r)$.

Precisamos provar a afirmação de que W é um subespaço de V. Desta forma, enunciamos:

Teorema 5.11. *Seja $W = G(v_1, v_2, \ldots, v_r)$, onde $v_1, v_2, \ldots, v_r$ são vetores de um espaço vetorial V. Temos que W é um subespaço de V.*

Demonstração. Sejam $w_1, w_2 \in W$ e $\alpha \in F$. Temos:

$$w_1 = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_r v_r$$

e

$$w_2 = \beta_1 v_1 + \beta_2 v_2 + \cdots + \beta_r v_r,$$

onde $\alpha_i, \beta_i \in F$, $\forall i = \{1, \ldots, r\}$. Portanto,

$$\begin{aligned} w_1 + \alpha w_2 &= (\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_r v_r) + \alpha (\beta_1 v_1 + \beta_2 v_2 + \cdots + \beta_r v_r) \\ &= (\alpha_1 + \alpha\beta_1) v_1 + (\alpha_2 + \alpha\beta_2) v_2 + \cdots + (\alpha_r + \alpha\beta_r) v_r. \end{aligned}$$

Assim, $w_1 + \alpha w_2$ é uma combinação linear de $v_1, v_2, \ldots, v_r$ e consequentemente pertence a W. Pelo Teorema 5.4, W é um subespaço de V. □

Voltemos mais uma vez ao Exemplo 5.5. Vamos determinar um conjunto gerador para os subespaços U e W.

Exemplo 5.6. Relembrando, temos $U = \{(x, y) \in \mathbb{R}^2;\ x + y = 0\}$ e $W = \{(x, y) \in \mathbb{R}^2;\ x - y = 0\}$. Também vimos que $U = \{(-y, y) \in \mathbb{R}^2;\ y \in \mathbb{R}\}$ e $W = \{(x, x) \in \mathbb{R}^2;\ x \in \mathbb{R}\}$. Deste modo temos, $u = (-y, y) = y(-1, 1)$ e $w = (x, x) = x(1, 1)$. E portanto $U = G((-1, 1))$ e $W = G((1, 1))$.

Exemplo 5.7. Vamos determinar vetores que geram o espaço $W = \{(x, y, z) \in \mathbb{R}^3;\ x + 2y - 3z = 0\}$. Temos que:

$$\begin{aligned} (x, y, z) \in W &\iff x + 2y - 3z = 0 \\ &\iff x = -2y + 3z. \end{aligned}$$

Assim podemos escrever:

$$(x, y, z) = (-2y + 3z, y, z) = (-2y, y, 0) + (3z, 0, z) = y(-2, 1, 0) + z(3, 0, 1).$$

Portanto, temos que $(-2, 1, 0)$ e $(3, 0, 1)$ são vetores que geram o espaço W e assim podemos dizer que $G\left((-2, 1, 0), (3, 0, 1)\right)$ é um espaço gerado por W.

Para gerarmos um mesmo espaço vetorial, podemos usar conjuntos geradores distintos. Note que para gerarmos o $\mathbb{R}^2$ é usual usarmos $G\left((1, 0), (0, 1)\right)$. Entretanto, como vimos no Exercício 5.5, $\mathbb{R}^2 = U + W$, então pelo Exemplo 5.6 temos $G\left((-1, 1), (1, 1)\right)$ também gera $\mathbb{R}^2$.

O Teorema 5.12 que enunciaremos a seguir, nos traz uma condição necessária e suficiente para que conjuntos distintos de vetores gerem o mesmo espaço.

Teorema 5.12. *Considerem $\mathcal{A} = \{u_1, u_2, \ldots, u_r\}$ e $\mathcal{B} = \{w_1, w_2, \ldots, w_s\}$ dois conjuntos de vetores em um espaço vetorial V. Teremos que $G(u_1, u_2, \ldots, u_r) = G(w_1, w_2, \ldots, w_s)$ se, e somente se, cada vetor em $\mathcal{A}$ é uma combinação linear dos vetores de $\mathcal{B}$ e cada vetor em $\mathcal{B}$ é uma combinação linear dos vetores de $\mathcal{A}$.*

Demonstração. Suponhamos que $G(u_1, u_2, \ldots, u_r) = G(w_1, w_2, \ldots, w_s)$, gerem o espaço vetorial V. Assim dados $\alpha_i, \beta_j \in F$ com $i \in \{1, 2, \ldots, r\}$ e $j \in \{1, 2, \ldots, s\}$, podemos escrever $v \in V$ das seguintes formas:

$$v = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_r u_r \tag{5.1}$$

e

$$v = \beta_1 w_1 + \beta_2 w_2 + \cdots + \beta_s w_s. \tag{5.2}$$

Tomemos $\alpha_1 \neq 0$ e $\alpha_i = 0$ para todo $i \neq 1$ e substituímos em (5.1), desta forma temos:

$$v = \alpha_1 u_1. \tag{5.3}$$

Agora escrevendo v como combinação linear dos vetores do conjunto $\mathcal{B}$ conforme (5.2) e substituindo em (5.3), temos:

$$\beta_1 w_1 + \beta_2 w_2 + \cdots + \beta_s w_s = \alpha_1 u_1. \tag{5.4}$$

Por fim, multiplicamos a equação (5.4) por α_1^{-1}, deste modo:

$$\begin{array}{ccccccccccc}
\underbrace{\alpha_1^{-1}\beta_1}_{\gamma_{11}} w_1 & + & \underbrace{\alpha_1^{-1}\beta_2}_{\gamma_{12}} w_2 & + & \cdots & + & \underbrace{\alpha_1^{-1}\beta_s}_{\gamma_{1s}} w_s & = & \alpha_1^{-1}\alpha_1 u_1 & = & u_1 \\
\gamma_{11} w_1 & + & \gamma_{12} w_2 & + & \cdots & + & \gamma_{1s} w_s & = & u_1. & &
\end{array}$$

E assim escrevemos o vetor u_1 em $\mathcal{A}$, como uma combinação linear dos vetores de $\mathcal{B}$. Note que os índices i, j de $\gamma_{ij} \in F$ são organizados do seguinte modo, $i \in \{1, 2, \ldots, r\}$ e $j \in \{1, 2, \ldots, s\}$.

Fazendo o mesmo procedimento para todos os r vetores de $\mathcal{A}$, ou seja, tomando $\alpha_1 \neq 0$ e $\alpha_i = 0$ para todo $i \neq 1$, $\alpha_2 \neq 0$ e $\alpha_i = 0$ para todo $i \neq 2$, ..., $\alpha_r \neq 0$ e $\alpha_i = 0$ para todo $i \neq r$, obtemos as seguintes equações:

$$\begin{array}{ccccccccc}
\gamma_{11} w_1 & + & \gamma_{12} w_2 & + & \cdots & + & \gamma_{1s} w_s & = & u_1 \\
\gamma_{21} w_1 & + & \gamma_{22} w_2 & + & \cdots & + & \gamma_{2s} w_s & = & u_2 \\
\vdots & & \vdots & & \ddots & & \vdots & & \vdots \\
\gamma_{r1} w_1 & + & \gamma_{r2} w_2 & + & \cdots & + & \gamma_{rs} w_s & = & u_r.
\end{array}$$

Analogamente, podemos tomar $\beta_1 \neq 0$ e $\beta_j = 0$ para todo $j \neq 1$ e substituir na equação (5.2), encontrando $v = \beta_1 w_1$. Escrevendo v como combinação linear dos vetores do conjunto $\mathcal{A}$, já multiplicando todos os membros por β_1^{-1}, chegamos em:

$$\underbrace{\beta_1^{-1}\alpha_1}_{\delta_{11}} u_1 + \underbrace{\beta_1^{-1}\alpha_2}_{\delta_{12}} u_2 + \cdots + \underbrace{\beta_1^{-1}\alpha_r}_{\delta_{1r}} u_r = \beta_1^{-1}\beta_1 w_1 = w_1.$$

Agora fazendo o mesmo procedimento para todos os s vetores de $\mathcal{B}$, obtemos:

$$\begin{array}{ccccccccc}
\delta_{11} u_1 & + & \delta_{12} u_2 & + & \cdots & + & \delta_{1r} u_r & = & w_1 \\
\delta_{21} u_1 & + & \delta_{22} u_2 & + & \cdots & + & \delta_{2r} u_r & = & w_2 \\
\vdots & & \vdots & & \ddots & & \vdots & & \vdots \\
\delta_{s1} u_1 & + & \delta_{s2} u_2 & + & \cdots & + & \delta_{sr} u_r & = & w_s.
\end{array}$$

Reciprocamente, se temos um conjunto de vetores $\mathcal{A} = \{u_1, u_2, \ldots, u_r\}$, onde cada vetor em $\mathcal{A}$ pode ser escrito como uma combinação linear dos vetores do conjunto $\mathcal{B} = \{w_1, w_2, \ldots, w_s\}$ então o espaço gerado por $G(u_1, u_2, \ldots, u_r)$, pode ser reescrito de outra maneira. Para facilitar a notação, vamos tomar:

$$\left.\begin{array}{ccccccccc}
\gamma_{11} w_1 & + & \gamma_{12} w_2 & + & \cdots & + & \gamma_{1s} w_s & = & u_1 \\
\gamma_{21} w_1 & + & \gamma_{22} w_2 & + & \cdots & + & \gamma_{2s} w_s & = & u_2 \\
\vdots & & \vdots & & \ddots & & \vdots & & \vdots \\
\gamma_{r1} w_1 & + & \gamma_{r2} w_2 & + & \cdots & + & \gamma_{rs} w_s & = & u_r
\end{array}\right\} \iff u_i = \textstyle\sum_{j=1}^{s} \gamma_{ij} w_j,\ i = 1, \ldots, r$$

e com isto, podemos escrever $G(u_1, u_2, \ldots, u_r) = G(u_i) = G\left(\sum_{j=1}^{s} \gamma_{ij} w_j\right)$. Logo, conseguimos escrever o espaço gerado por $G(u_i)$ usando somente vetores do conjunto $\mathcal{B}$ e portanto $G(u_i) = G(w_j)$. □

5.4 HOMOMORFISMO

Podemos relacionar o Exemplo 5.3 com o Exemplo 5.1? Esta é uma pergunta natural, já que V_n é da forma $\alpha_0 + \alpha_1 x + \cdots + \alpha_{n-1} x^{n-1}$, onde $\alpha_i \in F$ e $F^{(n)}$ é da forma $(\alpha_0, \alpha_1, \ldots, \alpha_{n-1})$. Podemos imaginar que possamos levar os elementos de V_n em $F^{(n)}$, ou seja, V_n e $F^{(n)}$ sejam isomorfos como espaços vetoriais.

Assim como fizemos anteriormente, iremos definir homomorfismo, monomorfismo e isomorfismo, agora para espaços vetoriais, iniciando pela definição de homomorfismo.

Definição 5.13. Se U e V são espaços vetoriais sobre F, então a aplicação linear T, também chamada de *transformação linear*, de U em V é dita um *homomorfismo* se:

1. $T(u_1 + u_2) = T(u_1) + T(u_2)$,
2. $T(\alpha u_1) = \alpha T(u_1)$,

para todos $u_1, u_2 \in U$ e todo $\alpha \in F$.

Assim como vimos anteriormente, um homomorfismo é uma aplicação que preserva toda a estrutura básica de nosso sistema.

Se T, for injetiva, ela é chamada de *monomorfismo* e caso ela seja bijetiva ela é chamada de *isomorfismo*. Também vamos definir o núcleo da transformação linear T [1] de maneira semelhante a que fizemos antes, assim:

Definição 5.14. Seja T uma transformação linear de U em V. O núcleo de T, é o conjunto de vetores de U que são levados por T no vetor nulo de V, ou seja, $\{u \in U;\ T(u) = 0\}$.

1 Alguns livros denotam o núcleo da transformação linear T por **Ker**T ou **Nuc**T.

Notemos que o núcleo de T é um subconjunto não vazio de V , já que T leva o vetor nulo de U no vetor nulo de V. Mais ainda, o núcleo de T é um subespaço de V . Com efeito, se $v_1, v_2 \in \{u \in U;\ T(u) = 0\}$ e se $\alpha \in F$, então $v_1 + \alpha v_2 \in \{u \in U;\ T(u) = 0\}$, pois $T(v_1 + \alpha v_2) = T(v_1) + \alpha T(v_2) = 0 + \alpha 0 = 0$.

Teorema 5.15. *Seja T uma transformação linear de U em V. Temos que T é injetiva se, e somente se, o núcleo de T for igual ao vetor nulo.*

Demonstração. Se uma aplicação T é injetiva, então a equação $T(u) = 0$ só possui a solução $u = 0$. De fato, sendo T injetiva e como $T(0) = 0$, tem-se que $T(u) = 0 = T(0)$ implica que $u = 0$. Reciprocamente, suponhamos agora que o núcleo de T seja somente o vetor nulo. Tomemos u e v vetores em U. Se $T(u) = T(v)$, então $T(u) - T(v) = 0$. Equivalentemente, $T(u - v) = 0$. Assim, $u - v \in \{w \in U;\ T(w) = 0\}$. Como o núcleo de T é igual ao vetor nulo, segue-se que $u - v = 0$, logo $u = v$, mostrando a injetividade de T. □

O conjunto de todos os homomorfismos de U em V será indicado por Hom (U, V). Em particular temos Hom(U, F) que é o conjunto de todos os homomorfismos de U em F, onde F é um corpo e Hom(U, U) que é denominado o *anel das transformações lineares*, que é o conjunto de todos os homomorfismos de U no próprio U, este homomorfismo é chamado de *automorfismo*.

Voltemos agora aos Exemplos 5.1 e 5.3, para mostrar que de fato os espaços vetoriais formados nestes exemplos são isomorfos.

Exemplo 5.8. Sejam $U = F^{(n)}$ e $V = V_n$ espaços vetoriais. Vamos definir uma aplicação $T : U \to V$ para mostrar o isomorfismo entre $F^{(n)}$ e V_n. Para isto, vamos escrever $(\alpha_1, \alpha_2, \ldots, \alpha_n) \in F^{(n)}$ de modo que valha a igualdade $(\alpha_1, \alpha_2, \ldots, \alpha_n) = (\beta_0, \beta_1, \ldots, \beta_{n-1})$. Assim, basta tomarmos a aplicação T como:

$$T(\alpha_1, \alpha_2, \ldots, \alpha_n) = \beta_0 + \beta_1 x + \cdots + \beta_{n-1} x^{n-1}.$$

Sejam V um espaço vetorial sobre F e W um subespaço de V. Queremos construir espaços vetoriais formados por grupos quocientes V/W, de grupos abelianos. Sabemos que os elementos de V/W são as classes laterais $v + W$ onde $v \in V$. Como os grupos são abelianos, temos que V/W também é abeliano.

Temos que assegurar que as operações em V/W estejam bem definidas. É o que faremos. Sejam $\alpha \in F$ e $v + W \in V/W$, definamos $\alpha(v + W) = \alpha v + W$. Assim temos que

mostrar que, se $v + W = v' + W$ então $\alpha(v + W) = \alpha(v' + W)$. Para isto, notemos que $v + W = v' + W$, implica que $v - v' \in W$ e como W é um subespaço temos necessariamente $\alpha(v - v') \in W$. Usando o item 3 do Lema 5.2 podemos afirmar que $\alpha v - \alpha v' \in W$ e então $\alpha + W = \alpha v' + W$. Assim $\alpha(v + W) = \alpha v + W = \alpha v' + W = \alpha(v' + W)$ e portanto o produto está bem definido. Agora vamos mostrar que $v + W$ é um espaço vetorial. Já sabemos que V/W é um grupo abeliano com relação a operação +, assim devemos mostrar que os 4 axiomas são válidos. De fato, temos para todos $\alpha \in F$ e $v + W \in V/W$ está definido um elemento, indicado por $\alpha(v + W)$, em V/W, tal que:

1. $\alpha\left((v_1 + W) + (v_2 + W)\right) = \alpha(v_1 + W) + \alpha(v_2 + W) = \alpha v_1 + \alpha v_2 + W = \alpha(v_1 + v_2) + W$.
2. $(\alpha + \beta)(v + W) = \alpha(v + W) + \beta(v + W) = \alpha v + \beta v + W$.
3. $\alpha\left(\beta(v + W)\right) = (\alpha\beta)(v + W)$.
4. $1(v + W) = v + W$.

para todos $\alpha, \beta \in F$, $v_1 + W, v_2 + W \in V/W$ (onde 1 representa o elemento unidade de F com relação à multiplicação).

Lema 5.16. *Se V é um espaço vetorial sobre F e W é um subespaço de V, então V/W é um espaço vetorial sobre F, onde*

1. $(v_1 + W) + (v_2 + W) = (v_1 + v_2) + W$;
2. $\alpha(v_1 + W) = \alpha v_1 + W$;

para $v_1 + W, v_2 + W \in V/W$ e $\alpha \in F$.

5.5 DEPENDÊNCIA E INDEPENDÊNCIA LINEAR

Vimos na Seção 5.3, que um espaço vetorial V pode ser gerado por um por conjunto de vetores. Em geral, pode haver mais de uma maneira para se escrever um vetor em V como uma combinação linear de vetores do conjunto gerador deste espaço. Por exemplo, tomando o conjunto $\mathcal{A} = \{(1,1), (-1,1), (1,-1)\}$, podemos escrever $\mathbb{R}^2 = G\left((1,1), (-1,1), (1,-1)\right)$, assim temos:

$$(1,1) = 1(1,1) + 1(-1,1) + 1(1,-1).$$

Mas também podemos escrever:

$$(1,1) = 1(1,1) + 0(-1,1) + 0(1,-1).$$

Os vetores do conjunto $\mathcal{A}$ são ditos *linearmente dependentes* sobre F, pois podemos escrever o vetor nulo, usando elementos em F que não sejam todos nulos,

$$(0,0) = 0(1,1) + 1(-1,1) + 1(1,-1).$$

Definição 5.17. Sejam $v_1, v_2, \ldots, v_n$ vetores em espaço vetorial V. Dizemos que $v_1, v_2, \ldots, v_n$, são *linearmente dependentes* sobre F, se existem $\lambda_1, \lambda_2, \ldots, \lambda_n$ em F, não todos nulos, tais que $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n = 0$.

Caso não existam $\lambda_1, \lambda_2, \ldots, \lambda_n$ em F, não todos nulos, tais que $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n = 0$, dizemos que os vetores $v_1, v_2, \ldots, v_n$ vetores em espaço vetorial V, são *linearmente independentes* sobre F. Assim, se os vetores não são linearmente dependentes sobre F, ou simplesmente, linearmente dependentes, eles são ditos linearmente independentes. Equivalentemente, podemos dizer:

Definição 5.18. Sejam $v_1, v_2, \ldots, v_n$ vetores em espaço vetorial V. Dizemos que $v_1, v_2, \ldots, v_n$, são *linearmente independentes*, se a equação $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n = 0$ só é satisfeita, se $\lambda_1 = \lambda_2 = \cdots = \lambda_n = 0$.

Devemos estar atentos ao fato de que a dependência linear é função não só dos vetores, mas também do corpo. Por exemplo, o corpo dos números complexos é um espaço vetorial sobre o corpo dos números reais e também sobre o corpo dos números complexos. Repare que os vetores $v_1 = 1$ e $v_2 = i$ são linearmente independentes sobre os reais, contudo linearmente dependentes sobre os complexos. Uma vez que,

$$\lambda_1 v_1 + \lambda_2 v_2 = 0.$$

só é satisfeita para o corpo dos reais se $\lambda_1 = \lambda_2 = 0$, ao passo que para o corpo dos complexos a equação é satisfeita não somente pela solução trivial, mas também se $\lambda_1 = i$ e $\lambda_2 = -1$. Assim:

$$\lambda_1 v_1 + \lambda_2 v_2 = i(1) + (-1)i = 0.$$

Teorema 5.19. *Considere os vetores $v_1, \ldots, v_n$ em um espaço vetorial V, então eles são linearmente dependentes, se somente se, pelo menos um dos vetores pode ser escrito como uma combinação linear dos outros vetores.*

Demonstração. Suponhamos que $v_1, \ldots, v_n$ sejam linearmente dependentes. Então temos $\alpha_1, \ldots, \alpha_n \in F$, não todos nulos, tal que $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$ onde $\alpha_k \neq 0$, para algum K. Deste modo, isolando $\alpha_k v_k$ em $\alpha_1 v_1 + \cdots + \alpha_k v_k + \cdots + \alpha_n v_n = 0$, temos:

$$\begin{aligned}
\alpha_k v_k &= -\alpha_1 v_1 - \cdots - \alpha_{k-1} v_{k-1} - \alpha_{k+1} v_{k+1} - \cdots - \alpha_n v_n \\
\alpha_k^{-1} \alpha_k v_k &= \alpha_k^{-1} \left(-\alpha_1 v_1 - \cdots - \alpha_{k-1} v_{k-1} - \alpha_{k+1} v_{k+1} - \cdots - \alpha_n v_n \right) \\
\alpha_k^{-1} \alpha_k v_k &= \underbrace{-\alpha_k^{-1} \alpha_1}_{\beta_1} v_1 - \cdots \underbrace{-\alpha_k^{-1} \alpha_{k-1}}_{\beta_{k-1}} v_{k-1} \underbrace{-\alpha_k^{-1} \alpha_{k+1}}_{\beta_{k+1}} v_{k+1} - \cdots - \underbrace{\alpha_k^{-1} \alpha_n}_{\beta_n} v_n \\
v_k &= \beta_1 v_1 + \cdots + \beta_{k-1} v_{k-1} + \beta_{k+1} v_{k+1} + \cdots + \beta_n v_n.
\end{aligned}$$

E portanto v_k é uma combinação linear dos demais vetores. Reciprocamente, consideremos que exista um vetor v_i que possa ser escrito como combinação linear dos vetores $v_1, \ldots, v_n$. Então:

$$\begin{aligned}
v_i &= \gamma_1 v_1 + \cdots + \gamma_{i-1} v_{i-1} + \gamma_{i+1} v_{i+1} + \cdots + \gamma_n v_n \\
0 &= \gamma_1 v_1 + \cdots + \gamma_{i-1} v_{i-1} + 1 v_i + \gamma_{i+1} v_{i+1} + \cdots + \gamma_n v_n.
\end{aligned}$$

E assim, os vetores $v_1, \ldots, v_n$ são linearmente dependentes, uma vez que os $\gamma_1, \ldots, \gamma_n$ não são todos nulos, para $\gamma_1 v_1 + \cdots + \gamma_n v_n = 0$. □

Recordando a Definição 5.18, decorre do resultado demonstrado no Teorema 5.19 que, *dados vetores $v_1, \ldots, v_n$ em um espaço vetorial V, então eles são linearmente independentes, se somente se, nenhum dos vetores pode ser escrito como uma combinação linear dos outros vetores.*

5.6 BASE E DIMENSÃO

Vamos nesta seção introduzir dois conceitos muito importantes para os espaços vetoriais. Estes conceitos nos serão útil também no próximo capítulo. Como vimos na seção anterior, um conjunto de vetores que geram um espaço vetorial V, pode ser formado por vetores que são linearmente dependentes, ou linearmente independentes. Os conjuntos linearmente dependentes, de certa modo, aparenta ser um "desperdício",

por conter mais informação do que o necessário. E isso neste caso, não nos parece proveitoso. Seria mais interessante, se pudéssemos escrever cada vetor de modo único dentro de um espaço vetorial V como combinação linear dos vetores que geram este espaço vetorial V. O primeiro passo neste sentido é definirmos *base*.

Definição 5.20. Denominamos *base* um conjunto ordenado de vetores $\mathcal{B} = \{v_1, v_2, \ldots, v_n\}$ linearmente independentes de um espaço vetorial V, tais que o conjunto $\mathcal{B}$ gera V.

Vamos mostrar que da forma como definimos base, cada vetor do espaço vetorial V gerado por esta base é escrito de modo único.

Teorema 5.21. *Considere $\mathcal{B} = \{v_1, v_2, \ldots, v_n\}$ um conjunto ordenado de vetores do espaço vetorial V. Assim cada vetor v em V é escrito de maneira única como uma combinação linear dos vetores de $\mathcal{B}$, se e somente, se $\mathcal{B}$ for uma base.*

Demonstração. Suponhamos que $\mathcal{B}$ é uma base de V. Tomemos $v \in V$, assim existem $\alpha_1, \alpha_2, \ldots, \alpha_n \in F$ tais que

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n. \tag{5.5}$$

Suponhamos que existam $\beta_1, \beta_2, \ldots, \beta_n \in F$ tais que

$$v = \beta_1 v_1 + \beta_2 v_2 + \cdots + \beta_n v_n. \tag{5.6}$$

De (5.5) e (5.6), segue que:

$$(\alpha_1 - \beta_1)v_1 + (\alpha_2 - \beta_2)v_2 + \cdots + (\alpha_n - \beta_n)v_n = 0. \tag{5.7}$$

Como $\mathcal{B}$ é uma base, segue da Definição 5.20 que os $(\alpha_i - \beta_i) \in F$ na equação (5.7) são todos nulos, para todos $0 \leq i \leq n$. Assim $\alpha_i = \beta_i$ e portanto v é escrito de maneira única como uma combinação linear dos vetores de $\mathcal{B}$. Reciprocamente, se v é escrito de maneira única como uma combinação linear dos vetores de $\mathcal{B}$, então para $v = 0$, temos que

$$\gamma_1 v_1 + \gamma_2 v_2 + \cdots + \gamma_n v_n = 0,$$

implica que $\gamma_1, \gamma_2, \ldots, \gamma_n = 0$ e portanto $\mathcal{B}$ é uma base. □

Teorema 5.22. *Sejam $v_1, v_2, \ldots, v_n$ vetores não nulos que geram um espaço vetorial V. Então, dentre estes vetores, podemos extrair uma base de V.*

Demonstração. Consideremos $\mathcal{B}_0 = \{v_1, v_2, \ldots, v_n\}$. Se esses vetores geram uma base, não há o que demonstrar. Caso não gerem, isto implica que os vetores são linearmente dependentes, assim pelo Teorema 5.19, há um vetor entre eles que pode ser escrito como combinação linear dos outro vetores. Sem perda de generalidade, vamos supor que este vetor seja v_n, ou seja, v_n é uma combinação linear de $v_1, v_2, \ldots, v_{n-1}$. Assim podemos escrever $v_n = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_{n-1} v_{n-1}$ e portanto o conjunto $\mathcal{B}_1 = \{v_1, v_2, \ldots, v_{n-1}\}$ ainda gera V. Pois para qualquer vetor v em V, temos:

$$
\begin{aligned}
v &= \beta_1 v_1 + \beta_2 v_2 + \cdots + \beta_{n-1} v_{n-1} + \beta_n v_n \\
v &= \beta_1 v_1 + \beta_2 v_2 + \cdots + \beta_{n-1} v_{n-1} + \beta_n \left(\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_{n-1} v_{n-1}\right) \\
v &= \left(\beta_1 + \beta_n \alpha_1\right) v_1 + \left(\beta_2 + \beta_n \alpha_2\right) v_2 + \cdots + \left(\beta_{n-1} + \beta_n \alpha_{n-1}\right) v_{n-1} \\
v &= \gamma_1 v_1 + \gamma_2 v_2 + \cdots + \gamma_{n-1} v_{n-1}.
\end{aligned}
$$

Caso $\mathcal{B}_1$ seja linearmente independente, então $\mathcal{B}_1$ é uma base de V. Se $\mathcal{B}_1$ for linearmente dependente, então um dos vetores de $\mathcal{B}_1$ pode ser escrito como uma combinação linear dos outros vetores. Sem perda de generalidade, suponhamos v_{n-1} este vetor. Assim o conjunto $\mathcal{B}_2 = \{v_1, v_2, \ldots, v_{n-2}\}$ também gerará V. Se $\mathcal{B}_2$ é linearmente independente, então $\mathcal{B}_2$ é uma base, caso contrário, ou seja, se $\mathcal{B}_2$ for linearmente dependente prosseguimos como anteriormente. Após uma quantidade finita de passos, obteremos uma *base* que é o conjunto $\mathcal{B}_r$ formado por $n - r$ vetores ($0 \leq r \leq n - 1$) linearmente independentes que ainda geram V . □

Vamos introduzir o conceito de *dimensão*.

Definição 5.23. Um espaço vetorial não nulo V é chamado de *dimensão finita* se for gerado por uma base $\mathcal{B} = \{v_1, v_2, \ldots, v_n\}$. Se não existir tal conjunto, dizemos que o espaço tem *dimensão infinita*. Convencionamos que o espaço vetorial nulo é um espaço de dimensão finita.

Considere um espaço vetorial V, gerado pela base $\mathcal{B} = \{v_1, v_2, \ldots, v_n\}$. Dizemos que a *dimensão de* V é o número de vetores da base que gera V e denotamos por dim V. Convencionamos que se V é o espaço vetorial nulo, então dim $V = 0$.

Teorema 5.24. *Considere o espaço vetorial V de dimensão finita. Se $W \subseteq V$ com W subespaço de V, então W tem também dimensão finita e dim $W \leq$ dim V. Além disso, se dim $W =$ dim V , então $W = V$.*

Demonstração. Se $W = \{0\}$, W tem dimensão finita. Se $W \neq \{0\}$, tome $w_1 \in W$ com $w_1 \neq 0$. Assim o conjunto $\mathcal{B}_1 = \{w_1\}$ é linearmente independente. Se $\mathcal{B}_1$ gera W então $\mathcal{B}_1$ é uma base para W com dimensão finita igual a 1. Se $\mathcal{B}_1$ não gera W, então existe $w_2 \in W$ com $w_2 \notin G(w_1)$. O conjunto $\mathcal{B}_2 = \{w_1, w_2\}$ é linearmente independente. Desta maneira, se $\mathcal{B}_2$ gera W então $\mathcal{B}_2$ é uma base para W com dimensão finita igual a 2. Se $\mathcal{B}_2$ não gera W, prosseguimos com o raciocínio anterior. Como dim V é finita, digamos n, e qualquer conjunto linearmente independente de V tem no máximo n vetores, existe $m \in \mathbb{N}\backslash\{0\}$ com $m \leq n$ tal que

$$\mathcal{B}_m = \{w_1, w_2, \ldots, w_m\}$$

é uma base de W. Isto demonstra que W tem dimensão finita e dim $W = m$, com $m \leq n$.

Suponhamos agora que a dim W = dim $V = n$. Assim temos que uma base para W é $\mathcal{C} = \{w_1, w_2, \ldots, w_n\}$. Suponhamos também que $W \neq V$. Assim, existe $v \in V$ tal que $v \notin W$. Desta maneira o conjunto $\{w_1, w_2, \ldots, w_n, v\}$ é linearmente independente. Como este conjunto tem $n+1$ vetores a dim $V = n+1$, o que é uma contradição, uma vez que a dim $V = n$ e portanto $W = V$. □

6

RESOLUÇÃO POR RADICAIS

Ao longo dos Capítulos 4 e 5, foram apresentados vários conceitos e definições que mostram o quanto a álgebra evoluiu. Vimos que o tratamento algébrico dado aos conjuntos ampliou sobremaneira a percepção destes conjuntos. De certo modo, os números se mostram um entrave na evolução da resolução de certos problemas. Podemos citar o exemplo, de como, foi difícil para os primeiros matemáticos aceitarem o conjunto dos números complexos como sendo de fato um conjunto numérico.

Neste capítulo encerraremos o ciclo começado no Capítulo 4. Nosso objetivo é mostrar que não é possível resolver uma equação de grau 5, ou superior, utilizando uma fórmula, como fizemos nos primeiros capítulos. Entretanto, para alcançarmos nosso objetivo, bem mais foi mostrado pelo percurso. A matemática se interliga por vários caminhos e estes caminhos nos levam a lugares que muitas vezes, possuem vistas belíssimas que nos possibilita enxergar o horizonte de muito longe.

Um desses caminhos foi aberto por *Évariste Galois*,[1] que desenvolveu uma parte de grande relevância da álgebra abstrata, a teoria dos grupos. Isto entre muitas importantes contribuições dadas por ele.

6.1 EXTENSÃO DE CORPOS

Nesta seção vamos relacionar corpos. Lembrando que, corpo é um anel comutativo com elemento unidade no qual todo elemento não nulo possui um inverso multiplicativo, ou seja, um corpo é um anel comutativo no qual podemos dividir por qualquer elemento diferente de zero.

1 Foi um matemático francês, que nasceu em *Bourg-la-Reine*, 25 de outubro de 1811 e faleceu em *Paris*, 31 de maio de 1832, com 20 anos em um duelo.

Seja F um corpo, um corpo K é dito uma *extensão de* F se $K \supset F$. Em outras palavras, K é uma extensão de F se F é um subcorpo de K. Neste capítulo todo F indicará um dado corpo e K uma extensão de F.

Usaremos aqui os conceitos de dependência linear, dimensão, base, e outros definidos no Capítulo 5 de espaço vetorial, uma vez que, K é uma extensão de F, assim com as operações usuais de corpo em K, K é um espaço vetorial sobre F.

Definição 6.1. O *grau de K sobre F* é a dimensão de K como espaço vetorial sobre F e indicamos por $[K : F]$.

Quando K é de dimensão finita como espaço vetorial sobre F, dizemos que K é uma *extensão finita* de F. Isto nos permite enunciar o Teorema 6.2, que apesar de simples traz resultados significativos.

Teorema 6.2. *Se L é uma extensão finita de K e K é uma extensão finita de F, então L é uma extensão finita de F. Além disso, $[L : F] = [L : K]\,[K : F]$.*

Demonstração. Suponhamos que $\mathcal{B} = \{v_1, \ldots, v_m\}$ seja uma base de L sobre K e que $\mathcal{C} = \{w_1, \ldots, w_n\}$ seja uma base de K sobre F. Assim para $t \in L$ existem $k_1, \ldots, k_m$ em K, tal que:

$$t = k_1 v_1 + \cdots + k_m v_m. \tag{6.1}$$

Entretanto, cada $k_i \in K$ com $i \in \{1, \ldots, m\}$ pode ser escrito como combinação linear dos vetores de $\mathcal{C}$, assim:

$$\left.\begin{array}{ccccccccc} k_1 & = & f_{11}w_1 & + & f_{12}w_2 & + & \cdots & + & f_{1n}w_n \\ k_2 & = & f_{21}w_1 & + & f_{22}w_2 & + & \cdots & + & f_{2n}w_n \\ \vdots & & \vdots & & \vdots & & \ddots & & \vdots \\ k_m & = & f_{m1}w_1 & + & f_{m2}w_2 & + & \cdots & + & f_{mn}w_n \end{array}\right\} \iff k_i = \textstyle\sum_{j=1}^{n} f_{ij}w_j,\ i = 1, \ldots, m$$

com $f_{ij} \in F$.

Substituindo $k_i \in K$ em (6.1), temos:

$$t = \underbrace{\overbrace{\left(\sum_{j=1}^{n} f_{1j}w_j\right)}^{n \text{ parcelas}} v_1 + \cdots + \overbrace{\left(\sum_{j=1}^{n} f_{mj}w_j\right)}^{n \text{ parcelas}} v_m}_{m \text{ parcelas}}$$

Assim chegamos a

$$t = \sum_{i=1,j=1}^{m,n} f_{ij} v_i w_j.$$

Inicialmente supomos que $[L : K] = m$ e $[K : F] = n$ e com isso conseguimos expressar $t \in L$ como uma combinação linear sobre F dos elementos $v_i w_j$ com $f_{ij} \in F$. Agora, para mostrar que os vetores $v_i w_j$ formam uma base, ainda precisamos provar que estes vetores são linearmente independentes sobre F. Para isto, vamos supor que $\sum_{i=1,j=1}^{m,n} \alpha_{ij} v_i w_j = 0$. Assim, podemos escrever:

$$\begin{aligned}
\alpha_{11}v_1w_1 + \cdots + \alpha_{1n}v_1w_n + \cdots + \alpha_{m1}v_mw_1 + \cdots + \alpha_{mn}v_mw_n &= 0 \\
\left(\alpha_{11}w_1 + \cdots + \alpha_{1n}w_n\right) v_1 + \cdots + \left(\alpha_{i1}w_1 + \cdots + \alpha_{in}w_n\right) v_i + \left(\alpha_{m1}w_1 + \cdots + \alpha_{mn}w_n\right) v_m &= 0 \\
\underbrace{\left(\sum_{j=1}^{n} \alpha_{1j}w_j\right)}_{\tilde{k}_1} v_1 + \cdots + \underbrace{\left(\sum_{j=1}^{n} \alpha_{ij}w_j\right)}_{\tilde{k}_i} v_i + \cdots + \underbrace{\left(\sum_{j=1}^{n} \alpha_{mj}w_j\right)}_{\tilde{k}_m} v_m &= 0.
\end{aligned}$$

Note que $\alpha_{ij} \in K$, pois $K \supset F$. Tomando $\tilde{k}_i = \sum_{j=1}^{n} \alpha_{ij} w_j$, chegamos a $\tilde{k}_1 v_1 + \cdots + \tilde{k}_m v_m = 0$ e como os vetores $v_1, \ldots, v_m$ formam uma base sobre K, segue que $k_1 = \cdots = k_m = 0$. Por outro lado, $\tilde{k}_i = \sum_{j=1}^{n} \alpha_{ij} w_j$, para cada $i \in \{1, \ldots, m\}$. Como todos os $\tilde{k}_i = 0$ e os vetores $w_1, \ldots, w_n$ formam uma base de K sobre F, resulta que $\alpha_{ij} = 0$ e portanto os vetores $v_i w_j$ são linearmente independentes e formam uma base com mn elementos. Desta maneira concluímos que $[L : F] = mn = [L : K]\,[K : F]$. □

Corolário 6.3. *Se L é uma extensão finita de F e K é um subcorpo de L que contém F, então $[K : F] \mid [L : F]$.*

Demonstração. Pelo Teorema 6.2, temos que $[L : F] = [L : K]\,[K : F]$, assim

$$[L : K] = \frac{[L : F]}{[K : F]}$$

onde $[L : K] \in \mathbb{N}$ e portanto $[K : F] \mid [L : F]$. □

Definição 6.4. Um elemento $a \in K$ é dito *algébrico sobre F* se existem elementos $\alpha_0, \alpha_1, \ldots, \alpha_n$ em F, não todos nulos, tais que $\alpha_0 a^n + \alpha_1 a^{n-1} + \cdots + \alpha_n = 0$.

Dado $F[x]$ o anel dos polinômios em x sobre F. Se tomarmos $q(x) \in F[x]$, tal que $q(x) = \beta_0 x^m + \beta_1 x^{m-1} + \cdots + \beta_m$, então para todo elemento $b \in K$, indicamos por $q(b) = \beta_0 b^m + \beta_1 b^{m-1} + \cdots + \beta_m$ e dizemos que $q(b)$ é o *valor* do polinômio $q(x)$ quando se substitui x por b. Quando $q(b) = 0$ dizemos que b *satisfaz* $q(x)$. Nesta situação, $a \in K$ é algébrico sobre F se existe um polinômio não nulo $p(x) \in F[x]$ tal que $p(a) = 0$, ou seja, é satisfeito por a.

Vejamos um exemplo para ilustrar o que estamos tratando.

Exemplo 6.1. Seja $\mathbb{R}$ o corpo dos números reais e $\mathbb{Q}$ o corpo dos números racionais. Em $\mathbb{R}$, $\sqrt{2}$ e $\sqrt{3}$ são algébricos sobre $\mathbb{Q}$. Determinar um polinômio de grau 4 sobre $\mathbb{Q}$ satisfeito por $\sqrt{2}+\sqrt{3}$.

Resolução: Consideremos o polinômio de grau 4, $p(x) = e + dx + cx^2 + bx^3 + ax^4$, com $a, b, c, d, e \in \mathbb{Q}$. Queremos que $\left(\sqrt{2}+\sqrt{3}\right)$ seja raiz do polinômio $p(x)$, ou seja, $p\left(\sqrt{2}+\sqrt{3}\right) = 0$

Notemos inicialmente que:

$$\begin{aligned} x &= \left(\sqrt{2}+\sqrt{3}\right) \\ x^2 &= \left(\sqrt{2}+\sqrt{3}\right)^2 = 2\sqrt{6}+5 \\ x^3 &= \left(\sqrt{2}+\sqrt{3}\right)^3 = 9\sqrt{3}+11\sqrt{2} \\ x^4 &= \left(\sqrt{2}+\sqrt{3}\right)^4 = 20\sqrt{6}+49. \end{aligned}$$

Desta forma, substituindo a raiz $\left(\sqrt{2}+\sqrt{3}\right)$ no polinômio $e + dx + cx^2 + bx^3 + ax^4$ obtemos:

$$\begin{gathered} e + d\left(\sqrt{2}+\sqrt{3}\right) + c\left(2\sqrt{6}+5\right) + b\left(9\sqrt{3}+11\sqrt{2}\right) + a\left(20\sqrt{6}+49\right) = 0 \\ \sqrt{2}\,(d+11b) + \sqrt{3}\,(d+9b) + \sqrt{6}\,(2c+20a) + e + 5c + 49a = 0. \end{gathered}$$

Uma solução é tomarmos as multiplicações que envolvem $\sqrt{2}$, $\sqrt{3}$ e $\sqrt{6}$ iguais a zero, assim temos:

$$(d+11b) = 0 \tag{6.2}$$

$$(d+9b) = 0 \tag{6.3}$$

$$(2c+20a) = 0 \tag{6.4}$$

Igualando as equações (6.2) e (6.3) chegamos a $b = d = 0$. Da equação (6.4) chegamos a $c = -10a$.

Como devemos ter $p(x) = 0$ obrigatoriamente teremos $e + 5c + 49a = 0$. Tomando $a = 1$ (como um valor possível para a), obtemos $c = -10$ e assim:

$$\begin{aligned} e + 5c + 49a &= 0 \\ e + 5(-10) + 49(1) &= 0 \\ e &= 1. \end{aligned}$$

Assim obtemos o polinômio $x^4 - 10x^2 + 1$ que satisfaz as condições do enunciado.

É fácil fazer a verificação, uma vez que $x^4 - 10x^2 + 1 = x^2\left(x^2 - 10\right) + 1$ e tomando $x = \left(\sqrt{2} + \sqrt{3}\right)$ temos:

$$\begin{aligned} x^2\left(x^2 - 10\right) + 1 &= \left(2\sqrt{6} + 5\right)\left(2\sqrt{6} + 5 - 10\right) + 1 \\ &= \left(2\sqrt{6} + 5\right)\left(2\sqrt{6} - 5\right) + 1 \\ &= 4 \cdot 6 - 25 + 1 = 0 \end{aligned}$$

Agora tomemos um $a \in K$ onde K é uma extensão de F. Considere $\mathcal{M}$ a coleção de todos os subcorpos de K que contêm F e a. $\mathcal{M}$ não é vazio, pois certamente K é um elemento de $\mathcal{M}$. Como a interseção de um número qualquer de subcorpos de K é novamente um subcorpo de K, temos que a interseção de todos os subcorpos de K que são elementos de $\mathcal{M}$ é um subcorpo de K. Indicamos , este subcorpo por $F(a)$. Note que $F(a)$ contém F e a, pois $F(a)$ é subcorpo de K que é membro de $\mathcal{M}$. Além disso, pela própria definição de interseção, todo subcorpo de K em $\mathcal{M}$ contém $F(a)$ e ao mesmo tempo $F(a)$ está em $\mathcal{M}$. Assim, $F(a)$ é o menor subcorpo de K que contém F e a. O subcorpo indicado por $F(a)$ é obtido pela *adjunção* de a e F.

Uma extensão de um corpo F pela adjunção de um elemento a algébrico em K, nos fornece obrigatoriamente uma extensão finita $F(a)$ de F. Isto está enunciado no teorema a seguir, que está demonstrado em [6].

Teorema 6.5. *O elemento $a \in K$ é algébrico sobre F se, e somente se, $F(a)$ é uma extensão finita de F.*

Definindo grau para um elemento algébrico nos permitirá estender um pouco mais o Teorema 6.5.

Definição 6.6. O elemento $a \in K$ é dito *algébrico de grau n* sobre F se ele satisfaz um polinômio não nulo sobre F de grau n, mas não satisfaz nenhum polinômio não nulo de grau menor.

Teorema 6.7. *Se $a \in K$ é algébrico de grau n sobre F, então $[F(a) : F] = n$.*

Voltemos a mais alguns exemplos.

Exemplo 6.2. Qual é o grau de $\sqrt{2} + \sqrt{3}$ sobre $\mathbb{Q}$?

Resolução: Como vimos no Exemplo 6.1 ao operarmos o número da forma $\sqrt{2} + \sqrt{3}$, obtemos números que "contenham" $\sqrt{2}$, $\sqrt{3}$ e $\sqrt{6}$. Desta maneira, olhando o conjunto $\mathbb{Q}\left(\sqrt{2}+\sqrt{3}\right)$ como um espaço vetorial, uma base para ele é o conjunto $\mathcal{B} = \left\{1, \sqrt{2}, \sqrt{3}, \sqrt{6}\right\}$. Assim $\dim \mathbb{Q}\left(\sqrt{2}+\sqrt{3}\right) = 4$, logo $\left[\mathbb{Q}\left(\sqrt{2}+\sqrt{3}\right) : \mathbb{Q}\right] = 4$.

Exemplo 6.3. Qual é o grau de $\sqrt{2}\sqrt{3}$ sobre $\mathbb{Q}$?

Resolução: Queremos saber $\left[\mathbb{Q}\left(\sqrt{6}\right) : \mathbb{Q}\right]$, analogamente ao exercício anterior podemos tomar $\mathcal{C} = \left\{1, \sqrt{6}\right\}$ como base do espaço vetorial de $\mathbb{Q}\left(\sqrt{6}\right)$, assim $\dim \mathbb{Q}\left(\sqrt{6}\right) = 2$. Logo $\left[\mathbb{Q}\left(\sqrt{6}\right) : \mathbb{Q}\right] = 2$.

Exemplo 6.4. Com a mesma notação do Exemplo 6.3, mostrar que $\sqrt{2} + \sqrt[3]{5}$ é algébrico de grau 6 sobre $\mathbb{Q}$.

Resolução: Desenvolvendo $\left(\sqrt{2} + \sqrt[3]{5}\right)$ obtemos os termos $\sqrt{2}, \sqrt[3]{5}, \sqrt{2}\sqrt[3]{5}, \sqrt[3]{25}$ e $\sqrt{2}\sqrt[3]{25}$. Assim $\mathcal{B} = \{1, \sqrt{2}, \sqrt[3]{5}, \sqrt{2}\sqrt[3]{5}, \sqrt[3]{25}, \sqrt{2}\sqrt[3]{25}\}$ é uma base para o espaço vetorial $\mathbb{Q}\left(\sqrt{2} + \sqrt[3]{5}\right)$, logo $\dim \mathbb{Q}\left(\sqrt{2} + \sqrt[3]{5}\right) = 6$ e portanto $\left[\mathbb{Q}\left(\sqrt{2} + \sqrt[3]{5}\right) : \mathbb{Q}\right] = 6$.

Se nos restringirmos aos conjuntos numéricos, podemos nos perguntar quais números são algébricos? Assim, vamos definir primeiro o que é um *número algébrico*.

Definição 6.8. Um número complexo é dito um *número algébrico* se ele é algébrico sobre o corpo dos números racionais.

Afirmamos que existem números complexos que não são algébricos. Estes números são chamados de *transcendentes*. E mais, os transcendentes formam um conjunto infinito, como vimos no primeiro capítulo, os conjuntos infinitos não têm todos o mesmo

"tamanho". Conforme mencionamos, Georg Cantor foi o responsável por mostrar uma relação entre conjuntos infinitos. Assim, pela Definição 1.4, o conjunto dos números algébricos têm a mesma cardinalidade dos números racionais, mas diferente do conjunto dos números transcendentes. O conjunto dos números algébricos é dito *enumerável*, enquanto o dos números transcendentes é dito *não enumerável*. Podemos dar dois exemplos de números transcendentes bem conhecidos, o e e o π.

6.2 RAÍZES DE POLINÔMIOS

Nesta seção iremos buscar um corpo K que é uma extensão de F, tal que dado um polinômio $p(x)$ em $F[x]$, $p(x)$ tenha *raiz*. Assim, inicialmente definimos raiz.

Definição 6.9. Se $p(x) \in F[x]$, então um elemento a, que esteja em alguma extensão do corpo F, é denominado uma *raiz* de $p(x)$ se $p(a) = 0$.

Teorema 6.10. *do Resto*
Se $p(x) \in F[x]$ e K é uma extensão de F, então, para todo elemento $b \in K$, $p(x) = (x-b)q(x) + p(b)$, onde $q(x) \in K[x]$ e onde gr $q(x)$ = gr $p(x) - 1$.

Demonstração. Como $F \subset K$ temos então $p(x) \in K[x]$. Pelo algoritmo da divisão em $K[x]$ temos:

$$p(x) = (x-b)q(x) + r \text{ onde } q(x) \in K[x] \text{ e onde } r = 0 \text{ ou gr } r < \text{gr } (x-b) = 1.$$

Logo temos $r = 0$ ou gr $r = 0$ com $r \in K$. Quando $r = 0$, temos $p(x) = (x-b)q(x)$. Sabemos que gr $(x-b) = 1$ e portanto gr $p(x)$ =gr $q(x) + 1$ logo gr $q(x)$ = gr $p(x) - 1$. Analogamente se o gr $r = 0$, chegamos ao mesmo resultado. □

Corolário 6.11. *Se $a \in K$ é uma raiz de $p(x) \in F[x]$, onde $F \subset K$, então em $K[x]$, $(x-a) \mid p(x)$.*

Demonstração. Pelo Teorema 6.10, temos em $K[x]$, $p(x) = (x-a)q(x) + p(a) = (x-a)q(x)$, pois $p(a) = 0$, logo $(x-a) \mid p(x)$. □

Vamos agora, contar quantas raízes um polinômio $p(x)$ em um corpo K possui. Primeiro vamos definir *raiz de multiplicidade m*, para que possamos fazer a contagem.

Definição 6.12. O elemento $a \in K$ é uma raiz de $p(x) \in F[x]$ de multiplicidade m se $(x-a)^m \mid p(x)$, enquanto $(x-a)^{m+1} \nmid p(x)$.

Com esta definição estamos prontos para contarmos as raízes e mostrarmos que:

Lema 6.13. *Um polinômio de grau n sobre um corpo pode ter no máximo n raízes em qualquer extensão deste corpo.*

Demonstração. Vamos provar por indução sobre n. Para $n = 1$, temos um polinômio da forma $p(x) = \alpha x + \beta$ com $\alpha \neq 0$, onde a única raiz de $p(x)$ é claramente $-\frac{\beta}{\alpha}$, pois $p\left(-\frac{\beta}{\alpha}\right) = \alpha\left(-\frac{\beta}{\alpha}\right) + \beta = 0$. Assim o lema é válido para o caso $n = 1$. Agora, vamos supor que o lema é válido para um polinômio de grau n e mostrar que esta validade acarretará na validade do lema para um polinômio de grau $n+1$. Primeiro observamos que se $p(x)$ não possuir raízes o lema está demonstrado. Agora, suponhamos que $p(x)$ possua pelo menos uma raiz $a \in K$ e que a seja de multiplicidade m. Como $(x-a)^m \mid p(x)$, temos que $m \leq n$. Ora $p(x) = (x-a)^m q(x)$, onde $q(x) \in K[x]$ e $\operatorname{gr} q(x) = n - m$. Com efeito, $(x-a)^{m+1} \nmid p(x)$ implica que $(x-a) \nmid q(x)$. Logo pelo Corolário 6.11, a não é raiz de $q(x)$. Se $b \neq a$ é uma raiz em K de $p(x)$, então $0 = p(b) = (b-a)^m q(b)$, como $(b-a)^m \neq 0$ temos $q(x) = 0$, isto é, qualquer raiz de $p(x)$ em K, diferente de a é necessariamente uma raiz de $q(x)$. Como $\operatorname{gr} q(x) = n - m < n$, pela nossa hipótese de indução, $q(x)$ têm no máximo $n - m$ raízes em K. Desta forma, o número de raízes em $p(x)$ é no máximo as m raízes a mais do que as $n - m$ raízes em $q(x)$, logo temos $m + (n - m) = n$ raízes. □

Antes de apresentar o próximo teorema, vamos resolver dois problemas que envolve alguns conceitos já tratados que usaremos a seguir. Faremos um problema mais específico e um caso geral do primeiro problema dado.

Problema 2. Se $f(x) = x^2 + 1$ em $F[x]$, onde $F = \mathbb{Z}_3$, ou seja, o corpo dos inteiros mod 3. Mostre que $F[x]/\left(x^2+1\right)$, que é uma extensão de F onde $f(x)$ admite raiz, é um corpo com $3^2 = 9$ elementos.

Solução: Notemos que $A = \left(x^2+1\right)$ é o ideal de $F[x]$ gerado por $x^2 + 1$. Todo elemento $F[x]/\left(x^2+1\right)$ é uma classe lateral da forma $f(x) + A$ do ideal A, com $f(x)$ em

$F[x]$. Pelo algoritmo da divisão $f(x) = t(x)\left(x^2+1\right) + r(x)$, onde $r(x) = 0$ ou gr $r(x) < 2$. Assim, $r(x) = a_0 + a_1 x$, onde a_0 e a_1 estão em $F = \{[0], [1], [2]\}$. Consequentemente $f(x) + A = a_0 + a_1 x + t(x)\left(x^2+1\right) + A = a_0 + a_1 x + A = (a_0 + A) + a_1(x + A)$. Colocando $t = x + A$, temos que todo elemento $F[x]/\left(x^2+1\right)$ é da forma $a_0 + a_1 t$ com a_0 e a_1 em F. Como $t^2 + 1 = (x + A)^2 + 1 = x^2 + 1 + A = A = 0$, pois A é o elemento zero de $F[x]/\left(x^2+1\right)$, assim $t^2 = [2]$, pois $[2] + 1 = 0$.

Podemos verificar que o polinômio não possui nenhuma raiz em F. Note que:

$$\begin{aligned} [0]^2 + 1 &= [1] \\ [1]^2 + 1 &= [2] \\ [2]^2 + 1 &= [2]. \end{aligned}$$

E portanto (x^2+1) é irredutível sobre F. Pelo Lema 4.99, como (x^2+1) é irredutível sobre F, temos que (x^2+1) é um ideal maximal. E pelo Teorema 4.68, por (x^2+1) ser um ideal maximal de $F[x]/\left(x^2+1\right)$, temos que $F[x]/\left(x^2+1\right)$ é um corpo, pois $F[x]$ é um anel comutativo com unidade. E ainda mais, $F[x]/(x^2+1)$ é o menor corpo contendo $\mathbb{Z}_2$ onde x^2+1 tem raiz. Neste caso podemos mostrar diretamente que $F[x]/\left(x^2+1\right)$ é um corpo. Para isto devemos mostrar que se $a_0 + a_1 t \neq 0$ então ele possui um inverso da forma $\alpha + \beta t$ tal que $(a_0 + a_1 t)(\alpha + \beta t) = [1]$. Resolvendo para α e β e usando $t^2 = [2]$ temos: $(a_0 + a_1 t)(\alpha + \beta t) = a_0\alpha + a_0\beta t + a_1\alpha t + a_1\beta t^2 = a_0\alpha + a_0\beta t + a_1\alpha t + [2]\, a_1\beta = (a_0\alpha + [2]\, a_1\beta) + (a_0\beta + a_1\alpha)t = [1]$, assim:

$$\begin{aligned} a_0\alpha + [2]\, a_1\beta &= [1] \\ a_1\alpha + a_0\beta &= [0]. \end{aligned}$$

Resolvendo o sistema acima temos:

$$\begin{aligned} \alpha &= -\frac{a_0}{[2]\, {a_1}^2 - {a_0}^2} \\ \beta &= \frac{a_1}{[2]\, {a_1}^2 - {a_0}^2}. \end{aligned}$$

Como $[2]\, {a_1}^2 - {a_0}^2 \neq 0$ para todos os a_0 e a_1 em F, temos que todos elementos de $F[x]/\left(x^2+1\right)$ possui inverso, logo é um corpo.

Quanto aos elementos de $F[x]/\left(x^2+1\right)$, já sabemos que é da forma $f(x)+A$, assim temos que $F[x]/\left(x^2+1\right) = \{[0], [1], [2], [1]\,t, [2]\,t, [1]+[1]\,t, [2]+[1]\,t, [1]+[2]\,t, [2]+[2]\,t\}$, ou seja, possui 9 elementos.

Problema 3. Se $f(x)$ está em $F[x]$, onde F é o corpo dos inteiros $\mod p$, p um primo, e $f(x)$ é irredutível sobre F e de grau n, demonstrar que $F[x]/(f(x))$ é um corpo com p^n elementos.

Solução: Procedendo de modo análogo ao problema anterior, vamos mostrar que $F[x]/(f(x))$ é um corpo com p^n elementos, sabendo que $f(x)$ está em $F[x]$, onde F é o corpo dos inteiros $\mod p$, p um primo, e $f(x)$ é irredutível sobre F e de grau n.

Pelo Lema 4.99 o ideal $A = (p(x))$ em $F[x]$ é um ideal maximal se, e somente se, $p(x)$ é irredutível sobre $F[x]$ e portanto $f(x)$ é um ideal maximal. E pelo Teorema 4.68, por $f(x)$ ser um ideal maximal de $F[x]/(f(x))$, temos que $F[x]/(f(x))$ é um corpo, pois $F[x]$ é um anel comutativo com unidade. E assim a primeira parte do problema está resolvida.

Agora sabendo que $A = (f(x))$ é um ideal maximal. Pelo algoritmo da divisão $m(x) = t(x)(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n) + r(x)$, onde $r(x) = 0$ ou gr $r(x) < n$. Assim $r(x) = a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1}$ com $a_0, a_1, \ldots, a_{n-1} \in F$. Todo elemento $F[x]/(f(x))$ é uma classe lateral da forma $m(x) + A = a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1} + A = (a_0 + A) + a_1(x + A) + a_2(x + A)^2 + \cdots + a_{n-1}(x + A)^{n-1}$ chamando $x + A = t$, obtemos $m(t) = a_0 + a_1t + a_2t^2 + \cdots + a_{n-1}t^{n-1}$. Como $a_0, a_1, \ldots, a_{n-1} \in F$ e $F = \{[1], [2], \ldots, [p]\}$ podemos escrever:

$$m(t) = \sum_{i=1,j=0}^{p,n} = [i]\,t^j.$$

Logo $F[x]/(f(x))$ possui p^n elementos.

Vamos agora apresentar um teorema que nos direciona num caminho para determinar extensões convenientes de F nas quais um polinômio $p(x)$ em $F[x]$ tenha raízes.

Teorema 6.14. *Se $p(x)$ é um polinômio em $F[x]$, de grau $n \geq 1$, e é irredutível sobre F, então existe uma extensão E de F, tal que $[E : F] = n$, na qual $p(x)$ tem uma raiz.*

Demonstração. Sejam $F[x]$ o anel dos polinômios em x sobre F e $V = (p(x))$ o ideal de $F[x]$ gerado por $p(x)$. Pelo Lema 4.99, V é um ideal maximal de $F[x]$, donde, pelo Teorema 4.68, $E = F[x]/V$ é um corpo. Demonstraremos que este F satisfaz as conclusões do teorema.

Primeiramente queremos mostrar que E é uma extensão de F; contudo, na verdade, não é! Mas seja $\overline{F}$ a imagem de F em E, isto é, $\overline{F} = \{\alpha + V; \alpha \in F\}$. Afirmamos que $\overline{F}$ é um corpo isomorfo a F, de fato, se ψ é a aplicação de $F[x]$ em $F[x]/V = E$ definida por $\psi(x) = x + V$, então a restrição de ψ a F induz um isomorfismo de F em $\overline{F}$. Notemos que para $\forall \alpha, \beta \in F$ temos $\psi(\alpha)\psi(\beta) = (\alpha + V)(\beta + V) = (\alpha\beta + V) = \psi(\alpha\beta)$, com $\alpha + V, \beta + V \in \overline{F}$ e $\forall \alpha, \beta \in F$ temos $\psi(\alpha) + \psi(\beta) = (\alpha + V) + (\beta + V) = (\alpha + \beta + V) = \psi(\alpha + \beta)$, com $\alpha + V, \beta + V \in \overline{F}$. Usando este isomorfismo identificamos F em $\overline{F}$; *desta maneira podemos considerar E como uma extensão de F.*

Afirmamos que E é uma extensão finita de F de grau $n = \text{gr } p(x)$, pois o conjunto $\mathcal{B} = \{1, x + V, x^2 + V, \ldots, x^{n-1} + V\}$ com n elementos, formam uma base de E sobre F. Isto porque todo elemento em $F[x]/(p(x))$ é uma classe lateral da forma $f(x) + V$ do ideal V, com $f(x)$ em $F[x]$. Ora, dado um polinômio qualquer $F[x]$, pelo algoritmo da divisão, $f(x) = t(x)p(x) + r(x)$, onde $r(x) = 0$ ou $\text{gr } r(x) < \text{gr } p(x) = n$. Assim, $r(x) = a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1}$, onde $a_0, a_1, a_2, \ldots, a_{n-1}$ estão em F; consequentemente, $f(x) + V = a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1} + t(x)p(x) + V = a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1} + V$ pois $t(x)p(x)$ está em V, donde pela adição e multiplicação em $F[x]/(p(x))$, $f(x) + V = (a_0 + V) + a_1(x + V) + a_2(x + V)^2 + \cdots + a_{n-1}(x + V)^{n-1}$. Podemos então ter como base o conjunto $\mathcal{B} = \{(1 + V), (x + V), (x + V)^2, \cdots, (x + V)^{n-1}\}$, como $(x + V)^2 = x^2 + V, \ldots, (x + V)^i = x^i + V, \ldots, (x + V)^{n-1} = x^{n-1} + V$, concluímos que $\mathcal{B} = \{1, x + V, x^2 + V, \ldots, x^{n-1} + V\}$ é uma base de E sobre F.

Agora nos falta mostrar que $p(x)$ tem uma raiz em E. Por simplicidade de notação indiquemos o elemento $x\psi = x + V$ no corpo E por a. Assim dado $f(x) \in F[x]$, com $f(x) = \beta_0 + \beta_1x + \cdots + \beta_kx^k$, temos $f(x)\psi = \beta_0\psi + (\beta_1\psi)(x\psi) + \cdots + (\beta_k\psi)(x\psi)^k$ e portanto $f(x)\psi = f(a)$. Em particular, como $p(x) \in V$, $p(x) = 0$; contudo, $p(x) = p(a)$. *Assim, o elemento $a = x\psi$ em E é uma raiz de $p(x)$.* Demonstramos que o corpo E satisfaz todas as propriedades requeridas na conclusão do Teorema 6.14 e assim concluímos a demonstração. □

Deste teorema podemos concluir:

Corolário 6.15. *Se $f(x) \in F[x]$, então existe uma extensão finita E de F na qual $f(x)$ tem uma raiz. Além disso, $[E : F] \leq gr\ f(x)$.*

Demonstração. Considere um polinômio $p(x)$ irredutível sobre F. Pelo Teorema 6.14 existe uma extensão E de F com $[E : F] = \text{gr } p(x)$. Agora tome $f(x) \in F[x]$ tal que $p(x)$ seja um fator irredutível de $f(x)$, assim $[E : F] = \text{gr } p(x) \leq \text{gr } f(x)$. Portanto, $f(x)$ tem uma raiz. □

Ainda podemos ir mais adiante sobre as extensões e chegar a mais um corolário.

Corolário 6.16. *do complemento completo de raízes*
Seja $f(x) \in F[x]$ de grau $n \geq 1$. Então, existe uma extensão E de F, de grau no máximo $n!$, na qual $f(x)$ possui n raízes.

Demonstração. No enunciado do teorema uma raiz de multiplicidade m é contada, obviamente, como m raízes.

Pelo Corolário 6.15 existe E_0 extensão de F tal que $[E_0 : F] \leq n$, na qual $f(x)$ possui uma raiz α. Ainda pelo mesmo corolário, temos que $f(x)$ pode ser fatorado, deste modo podemos escrever $f(x) = (x - \alpha)q(x)$, onde $q(x)$ é de grau $n - 1$ (caso base). Por sua vez, $q(x)$ também pode ser fatorado. Fazendo este processo (indutivo) um número finito de vezes, concluímos que $\text{gr } q(x) \leq (n-1)!$ e que existe uma extensão E de E_0, tal que $[E : E_0] \leq (n-1)!$, na qual $q(x)$ possui $n - 1$ raízes. Agora temos que $f(x) = (x - \alpha)q(x)$, possui a raiz α, mais as $n - 1$ raízes de $q(x)$, ou seja, n raízes. E para finalizar a demonstração, usando o Teorema 6.2, temos:

$$[E : F] = [E : E_0]\,[E_0 : F] \leq (n-1)!\, n = n!$$

□

O Corolário 6.16 afirma a existência de uma extensão finita E sobre F, na qual um polinômio $f(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n$, com $a_0 \neq 0$, possui n raízes. Se as n raízes em E são $\alpha_1, \ldots, \alpha_n$, usando o Corolário 6.11, $f(x)$ pode ser fatorado sobre E como $f(x) = a_0(x - \alpha_1)(x - \alpha_2)\cdots(x - \alpha_n)$, ou seja, $f(x)$ pode ser decomposta completamente sobre E como um produto de fatores *lineares* (de primeiro grau). Por outro lado, uma vez que existe uma extensão de F com esta propriedade, existe uma extensão finita de F de grau mínimo que também possui a propriedade de decomposição de $F(x)$ como produto de fatores lineares. Para esta extensão mínima, nenhum subcorpo próprio tem a propriedade de que $f(x)$ seja decomposto em fatores lineares sobre ele. E com isto definimos *corpo de raízes*.

Definição 6.17. Se $f(x) \in F[x]$, uma extensão finita E de F é dita um *corpo de raízes sobre F* para $f(x)$ se $f(x)$ pode ser expresso como um produto de fatores lineares sobre E, i.e., em $E[x]$, mas não sobre nenhum subcorpo próprio de E.

A Definição 6.17 é equivalente a dizer que E é um corpo de raízes de $f(x)$ sobre F se E é uma extensão mínima de F na qual possui n raízes, onde $n = \text{gr } f(x)$.

Vamos agora mostrar que existe um isomorfismo entre dois corpos de raízes, E_1 e E_2 do mesmo polinômio $f(x)$ em $F[x]$. Este isomorfismo deixa todo elemento de F fixo.

Sejam F e F' dois corpos e seja τ um isomorfismo de F em F'. Iremos, por simplicidade, indicar a imagem de qualquer α em F por intermédio de τ por α', i.e., $\tau(\alpha) = \alpha'$.

Lema 6.18. *Seja* $f(x) = \alpha_0 x^n + \alpha_1 x^{n-1} + \cdots + \alpha_n$ *um polinômio arbitrário em* $F[x]$. *Definamos* τ^* *um isomorfismo de* $F[x]$ *em* $F^{'}[t]$, *tal que* $\tau^*\left(f(x)\right) = \left(\alpha_0 x^n + \alpha_1 x^{n-1} + \cdots + \alpha_n\right) = \alpha_0^{'} t^n + \alpha_1^{'} t^{n-1} + \cdots + \alpha_n^{'}$. *Então* τ^* *goza da propriedade de que* $\tau^*(\alpha) = \alpha'$ *para todo* $\alpha \in F$.

Demonstração. Aplicando τ^* em $f(x)$, temos:

$$\begin{aligned}
\tau^*\left(f(x)\right) &= \left(\alpha_0 x^n + \alpha_1 x^{n-1} + \cdots + \alpha_n\right) \\
\tau^*\left(f(x)\right) &= \tau^*\left(\alpha_0 x^n\right) + \tau^*\left(\alpha_1 x^{n-1}\right) + \cdots + \tau^*\left(\alpha_n\right) \\
\tau^*\left(f(x)\right) &= \tau^*\left(\alpha_0\right)\tau^*\left(x^n\right) + \tau^*\left(\alpha_1\right)\tau^*\left(x^{n-1}\right) + \cdots + \tau^*\left(\alpha_n\right) \\
\tau^*\left(f(x)\right) &= \alpha_0^{'} t^n + \alpha_1^{'} t^{n-1} + \cdots + \alpha_n^{'}.
\end{aligned}$$

O que mostra a propriedade desejada. □

6.3 GRUPOS SOLÚVEIS

Nesta seção daremos, de certo modo, o primeiro passo em direção da nossa busca inicial, que é mostrar que um polinômio de grau $n \geq 5$ não é resolúvel por radicais. Para isto, vamos introduzir um grupo que terá papel de destaque para esta demonstração, o *grupo solúvel*.

Definição 6.19. *Grupo Solúvel*
Um grupo G é dito *solúvel* se pudermos determinar uma cadeia finitas de subgrupos $G = N_0 \supset N_1 \supset N_2 \supset \cdots \supset N_k = \{e\}$, onde cada N_i é um subgrupo normal de N_{i-1} e tal que todo grupo quociente N_{i-1}/N_i seja abeliano.

Antes de darmos um exemplo de grupo solúvel iremos mostrar um resultado que será útil em nosso exemplo.

Lema 6.20. *Todo grupo G com 2 ou 3 elementos é abeliano.*

Demonstração. Considere o grupo $G_1 = \{e, a\}$ onde e é o elemento unidade. Pela Definição 4.5 de grupos $ea = ae = a$. Logo este grupo é abeliano.

Considere agora o grupo $G_2 = \{e, a, b\}$. Vamos supor que este grupo não seja abeliano, assim $ab \neq ba$. Portanto b não é o inverso de a, pois se fosse $ab = ba = e$, logo o inverso de a é a. Como o grupo é fechado $ab = a$ ou $ab = b$. Vamos supor, sem perda de generalidade que $ab = a$ (assim $ba = b$). Assim temos

$$ab = a \iff aab = aa \iff eb = e \iff b = e,$$

o que é uma contradição. Logo G_2 é abeliano. □

Exemplo 6.5. O grupo simétrico de grau 3, S_3 é solúvel. Recordando que $S_3 = \{e, (1,2), (1,3), (2,3), (1,2,3), (1,3,2)\}$, onde e é o elemento unidade do grupo, que no caso do grupo de permutação, deixa todos os elementos fixos. Assim, usando a Definição 6.19, vou tomar $S_3 = N_0 \supset N_1 \supset N_2 = \{e\}$, com $N_1 = \{e, (1,2,3), (1,3,2)\}$. Usando a Definição 4.25, com o auxílio da Tabela 8, podemos comprovar que N_1 e N_2 são normais.

Evidentemente, $N_2 = \{e\}$ é normal, uma vez que pela definição N é um subgrupo normal de G se para todo $g \in G$ e $n \in N$, $gng^{-1} \in N$. Como N_2 consiste somente de e, temos que $geg^{-1} = gg^{-1}e = e \in N_2$.

Para N_1 usando a definição, devemos fazer a mesma verificação mencionada anteriormente. Faremos algumas. Observando a Tabela 8 para facilitar o processo, percebemos que e, $(1,2)$, $(1,3)$ e $(2,3)$ são inversos de si mesmos, enquanto $(1,2,3)$ é o

$\circ$	e	$(1,2,3)$	$(1,3,2)$	$(2,3)$	$(1,3)$	$(1,2)$
e	e	$(1,2,3)$	$(1,3,2)$	$(2,3)$	$(1,3)$	$(1,2)$
$(1,2,3)$	$(1,2,3)$	$(1,3,2)$	e	$(1,2)$	$(2,3)$	$(1,3)$
$(1,3,2)$	$(1,3,2)$	e	$(1,2,3)$	$(1,3)$	$(1,2)$	$(2,3)$
$(2,3)$	$(2,3)$	$(1,3)$	$(1,2)$	e	$(1,2,3)$	$(1,3,2)$
$(1,3)$	$(1,3)$	$(1,2)$	$(2,3)$	$(1,3,2)$	e	$(1,2,3)$
$(1,2)$	$(1,2)$	$(2,3)$	$(1,3)$	$(1,2,3)$	$(1,3,2)$	e

Tabela 8: Resultado das operações de composição entre elementos do grupo S_3.

inverso de $(1,3,2)$ e vice-versa. Assim, tomando $g, g^{-1} \in G$ e o elemento $(1,2,3) \in N_1$, obtemos $gng^{-1} \in N_1$:

$$\begin{aligned}(1,2)(1,2,3)(1,2) = (2,3)(1,2) = (1,3,2)\\ (1,3)(1,2,3)(1,3) = (1,2)(1,3) = (1,3,2)\\ (2,3)(1,2,3)(2,3) = (1,3)(2,3) = (1,3,2)\\ (1,2,3)(1,2,3)(1,3,2) = (1,3,2)(1,3,2) = (1,2,3)\\ (1,3,2)(1,2,3)(1,2,3) = e(1,2,3) = (1,2,3).\end{aligned}$$

O mesmo pode ser feito se tomarmos $(1,3,2) \in N_1$ e evidentemente para $e \in N_1$. Desta maneira, mostramos que N_1 e N_2 são subgrupos normais. Agora, note que:

$$o\left(S_3/N_1\right) = \frac{o(S_3)}{o(N_1)} = \frac{6}{3} = 2 \quad \text{e} \quad o\left(N_1/e\right) = \frac{o(N_1)}{o(e)} = \frac{3}{1} = 3$$

e portanto são abelianos.

Outro caminho é definir a solubilidade partindo da geração de um grupo abeliano. Desta maneira, diremos que G' é o *subgrupo comutador* de G, gerado pelo *comutador* $a^{-1}b^{-1}ab$ de a e b, onde $a, b \in G$. Assim G' é um subgrupo comutador de G gerado por todos comutadores em G. (Não é verdadeiro que o conjunto dos comutadores seja um subgrupo de G necessariamente.) Podemos observar que da maneira que definimos G', G/G' é abeliano. Pois, quaisquer que sejam aG' e bG', com $a, b \in G$, então

$$(aG')(bG') = abG' = eabG' = \left(baa^{-1}b^{-1}\right)abG' = ba\underbrace{\left(a^{-1}b^{-1}ab\right)}_{\in G'}G' = baG' = (bG')(aG').$$

Além disso, G' é subgrupo normal de G, pois, se $u \in G'$ e $g \in G$ temos:

$$gug^{-1} = (gug^{-1})u^{-1}u = (gug^{-1}u^{-1})u.$$

Daí, como $(gug^{-1}u^{-1}) \in G'$ e $u \in G'$, temos que $gug^{-1} \in G'$.

Finalmente, se M é um subgrupo normal de G tal que G/M seja abeliano, implica em $M \supset G'$, uma vez que, dados $a, b \in G$, acarreta em $(aM)(bM) = (bM)(aM)$ e assim concluímos que $abM = baM$, logo $a^{-1}b^{-1}abM = M$ e então $a^{-1}b^{-1}ab \in M$. Como M contém todos os comutadores, ele contém G' que é o grupo que estes geram.

G' é um grupo, assim podemos considerar seu subgrupo comutador $G^{(2)} = (G')'$. Então, neste caso $G^{(2)}$ é um subgrupo comutador de G' gerado por todos os elementos $(a')^{-1}(b')^{-1}a'b'$ de a' e b', onde $a', b' \in G'$. E também de modo análogo, M' é um subgrupo normal de G', tal que $M' \supset G^{(2)}$ com G'/M' abeliano.

Podemos afirmar que $G^{(2)}$ é um subgrupo normal também de G. Notemos que se M é um subgrupo normal de G tal que G/M seja abeliano, implica em $M \supset G'$. Mas M' é um subgrupo normal de G' tal que G'/M' é abeliano, assim $M' \supset G^{(2)}$. Logo, podemos tomar $M' = G'$, deste modo $M \supset G' \supset G^{(2)}$. E portanto, $G^{(2)}$ é um subgrupo normal de G. Seguindo deste modo, por indução, definimos os subgrupos comutadores $G^{(m)}$ por $G^{(m)} = \left(G^{(m-1)}\right)'$, onde $G^{(m)}$ é um subgrupo normal de G e $G^{(m-1)}/G^{(m)}$ é um grupo abeliano.

Partindo destes comutadores superiores de G, temos um critério bastante sucinto para a solubilidade de um grupo G.

Lema 6.21. *G é solúvel se, e somente se, $G^{(k)} = \{e\}$ para algum k inteiro.*

Demonstração. Se $G^{(k)} = \{e\}$ então podemos tomar $N_0 = G$, $N_1 = G'$, $N_2 = G^{(2)}$, ..., $N_k = G^{(k)} = \{e\}$, com cada N_i sendo normal em G, certamente será normal em N_{i-1}. E N_{i-1}/N_i será abeliano, pois:

$$\frac{N_{i-1}}{N_i} = \frac{G^{(i-1)}}{G^{(i)}} = \frac{G^{(i-1)}}{\left(G^{(i-1)}\right)'}.$$

Assim, pela Definição 6.19, G é um grupo solúvel. Reciprocamente, se G é solúvel, então temos $G = N_0 \supset N_1 \supset N_2 \supset \cdots \supset N_k = \{e\}$, onde cada N_i é um subgrupo normal de N_{i-1} com N_{i-1}/N_i abeliano. Deste modo, o subgrupo comutador N'_{i-1} de N_{i-1} está contido em N_i. Assim, $N_1 \supset N_0 = G'$, $N_2 \supset N'_1 \supset (G')' = G^{(2)}$, $N_3 \supset$

$N_2' \supset \left(G^{(2)}\right)' = G^{(3)}, \ldots, N_k \supset N_{k-1}' \supset \left(G^{(k-1)}\right)' = G^{(k)} = \{e\}$. Portanto, obtemos $G^{(k)} = \{e\}$. □

Corolário 6.22. *Se G é um grupo solúvel e $\overline{G}$ é uma imagem homomorfa de G, então $\overline{G}$ é solúvel.*

Demonstração. Como $\overline{G}$ é uma imagem homomorfa de G, segue que $\overline{G}^{(k)}$ é imagem de $G^{(k)}$. Deste modo $G^{(k)} = \{e\}$ para algum k e assim $\overline{G}^{(k)} = \{e\}$ para um mesmo k, logo pelo Lema 6.21 $\overline{G}$ é solúvel. □

Mostramos no Exemplo 6.5 que o grupo simétrico de grau 3 é solúvel. O Lema 6.23 que apresentamos a seguir, nos mostra que grupos simétricos S_n de grau $n \geq 5$ não são solúveis.

Lema 6.23. *Considere $G = S_n$ onde $n \geq 5$, então $G^{(k)}$, para $k = 1, 2, \ldots$, contém todo 3-ciclo de S_n.*

Demonstração. Primeiro vamos mostrar que para um grupo arbitrário G, se N é um subgrupo normal de G, então N' também é subgrupo normal de G. Como N é subgrupo normal de G, temos para todo $g \in G$ e todo $n \in N$, $gng^{-1} \in N$. Por outro lado, como N' é subgrupo normal em N, para todo $n \in N$ e todo $n' \in N'$, temos $nn'n^{-1} \in N'$. Para mostrarmos que N' é subgrupo normal de G, note que qualquer que seja $n' \in N' \subset N$, podemos escrever $n' = n'n'(n')^{-1}$. Assim

$$gn'g^{-1} = g\left(n'n'(n')^{-1}\right)g^{-1} = \underbrace{gn'}_{\in N}\, n' \underbrace{(n')^{-1}g^{-1}}_{\in N}.$$

Portanto $gn'g^{-1} \in N'$, logo N' é um subgrupo normal de G.

Afirmamos que se N é um subgrupo normal de $G = S_n$, onde $n \geq 5$, que contém todo 3-ciclo em S_n então N' também contém todo 3-ciclo.

Por hipótese $n \geq 5$, assim suponha $a, b \in N$, com $a = (1,2,3)$ e $b = (1,4,5)$, assim $a^{-1} = (3,2,1)$ e $b^{-1} = (5,4,1)$ e

$$\begin{aligned} a^{-1}b^{-1}ab &= (3,2,1)(5,4,1)(1,2,3)(1,4,5) \\ &= (3,2,5,4,1)(1,2,3)(1,4,5) \\ &= (2,5,4)(1,4,5) \\ &= (1,4,2). \end{aligned}$$

Como um comutador de elementos de N está necessariamente em N' e N' é um subgrupo normal de G, temos para todo $\rho \in S_n$, $\rho^{-1} \circ (1,4,2) \circ \rho$ também está em N'. Tomando $\rho \in S_n$ tal que $\rho(1) = i_1$, $\rho(4) = i_2$, $\rho(2) = i_3$, onde $i_1, i_2, i_3 \in \{1, 2, \ldots, n\}$ com $i_1 \neq i_2 \neq i_3$, sabendo que $\rho^{-1}(i_1) = 1$, $\rho^{-1}(i_2) = 4$ e $\rho^{-1}(i_3) = 2$, temos $\rho^{-1} \circ (1,4,2) \circ \rho = (i_1, i_2, i_3)$, pois $i_1 \to 1 \to 4 \to i_2$, $i_2 \to 4 \to 2 \to i_3$ e $i_3 \to 2 \to 1 \to i_1$. Como $(i_1, i_2, i_3) \in N'$, N' contém todos os 3-ciclos.

Colocando $N = G$, que é certamente normal em G e contém todos os 3-ciclos, obtemos que G' contém todos os 3-ciclos; como G' é normal em G, $G^{(2)}$ contém todos os 3-ciclos, como $G^{(2)}$ é normal em G, $G^{(3)}$ contém todos os 3-ciclos. Continuando desta maneira obtemos que $G^{(k)}$ contém todos os 3-ciclos para um k arbitrário. □

Como consequência direta do Lema 6.23 chegamos ao importante Teorema 6.24, que será uma das ferramentas utilizadas para mostrarmos que uma equação de grau $n \geq 5$ não possui fórmulas com radicais para determinar suas raízes.

Teorema 6.24. *O grupo simétrico S_n não é solúvel para $n \geq 5$.*

Demonstração. Vamos tomar $G = S_n$, com $n \geq 5$. Assim, pelo Lema 6.23, $G^{(k)}$ contém todos os 3-ciclos em S_n para todo k. Com isto, $G^{(k)} \neq \{e\}$ para todo k. Logo, S_n não pode ser solúvel, pois pelo Lema 6.21, um grupo só é solúvel se $G^{(k)} = \{e\}$.

□

6.4 GRUPOS DE GALOIS

No início da seção anterior, dissemos que o grupo solúvel poderia ser o primeiro passo na direção de nossa busca, em determinar que um polinômio de grau n com $n \geq 5$ não é resolúvel por radicais. Neste sentido e considerando que temos dois passos principais, o outro passo que daremos também será por intermédio de um grupo, o *grupo de Galois*.

Podemos estabelecer, como veremos adiante, uma relação entre o grupo de Galois e as raízes de um polinômio. E que o grupo de Galois se trata de um grupo de permutações das raízes do polinômio.

Considere um polinômio $p(x)$ em $F[x]$, o anel dos polinômios em x sobre F e associaremos a $p(x)$ um grupo. Este grupo é denominado *grupo de Galois* de $p(x)$.

Salientamos aqui, que consideraremos por hipótese os corpos, de agora adiante, com característica 0. Assim, estamos admitindo que os corpos dos polinômios que trataremos não têm raízes múltiplas. Neste sentido, nos é útil a seguinte definição:

Definição 6.25. A extensão K de E é uma *extensão simples* de F se $K = F(\alpha)$ para algum α em K.

Com isto, é demonstrado em [6] um teorema, que relaciona extensões algébricas de um corpo de característica 0. Enunciaremos sem demonstrar.

Teorema 6.26. *Se F é de característica 0 e a e b são algébricos sobre F, então existe um elemento $c \in F(a,b)$ tal que $F(a,b) = F(c)$.*

Exemplo 6.6. Para ilustrar o Teorema 6.26, tomemos $\sqrt{2}$ e $\sqrt{3}$ algébricos sobre $\mathbb{Q}$. Vamos mostrar que existe um elemento $\alpha \in \mathbb{Q}\left(\sqrt{2},\sqrt{3}\right)$ tal que $\mathbb{Q}\left(\sqrt{2},\sqrt{3}\right) = \mathbb{Q}(\alpha)$. Seja $\alpha = \sqrt{2}+\sqrt{3}$.

$$\alpha^2 = 2 + 2\sqrt{6} + 3 = 5 + 2\sqrt{6}$$
$$\alpha^3 = 5\sqrt{2} + 4\sqrt{3} + 5\sqrt{3} + 6\sqrt{2} = 11\sqrt{2} + 9\sqrt{3}.$$

Logo

$$\sqrt{2} = \frac{\alpha^3 - 9\alpha}{2} \qquad \text{e} \qquad \sqrt{3} = \frac{11\alpha - \alpha^3}{2}.$$

Logo $\mathbb{Q}\left(\sqrt{2}+\sqrt{3}\right) \supset \mathbb{Q}\left(\sqrt{2},\sqrt{3}\right)$. Mas obviamente $\mathbb{Q}\left(\sqrt{2}+\sqrt{3}\right) \subset \mathbb{Q}\left(\sqrt{2},\sqrt{3}\right)$.

Usando um argumento simples de indução, podemos mostrar que partindo de 2 elementos é possível ampliar este resultado para qualquer número finito, isto é, se $\alpha_1, \ldots, \alpha_n$ são algébricos sobre F, então existe um elemento $c \in F(\alpha_1, \ldots, \alpha_n)$, tal que $F(c) = F(\alpha_1, \ldots, \alpha_n)$. Assim:

Corolário 6.27. *Toda extensão finita de um corpo de característica 0 é uma extensão simples.*

Conforme propomos inicialmente, o grupo de Galois é uma associação de um polinômio $p(x)$ em $F[x]$, o anel dos polinômios em x sobre F e F um grupo. Para isto, faremos uso do corpo das raízes de $p(x)$ sobre F, assim o grupo de Galois pode ser também definido como um certo grupo de automorfismos deste corpo de raízes.

Aqui, estaremos tratando de *automorfismo do corpo K*, isto é uma função σ de K sobre si mesmo tal que $\sigma(a+b) = \sigma(a) + \sigma(b)$ e $\sigma(ab) = \sigma(a)\sigma(b)$ para todos $a, b \in K$. Dois automorfismos σ e τ de K são distintos se $\sigma(a) \neq \tau(a)$ para algum elemento $a \in K$.

Teorema 6.28. *Se K é um corpo e $\sigma_1, \ldots, \sigma_n$ são automorfismos distintos em K, então é impossível encontrar em K elementos $a_1, \ldots, a_n$, não todos nulos, tais que $a_1\sigma_1(w) + a_2\sigma_2(w) + \cdots + a_n\sigma_n(w) = 0$ para todo $w \in K$.*

Demonstração. Suponhamos por absurdo que exista $a_1, \ldots, a_n \in K$, não todos nulos, tais que $a_1\sigma_1(w) + a_2\sigma_2(w) + \cdots + a_n\sigma_n(w) = 0$ para todo $w \in K$. Assim, podemos tomar uma quantidade m mínima de $a_1, \ldots, a_n \in K$ tal que $a_1\sigma_1(w) + \cdots + a_m\sigma_m(w) = 0$, com $a_1, \ldots, a_m$ todos diferentes de 0.

Vamos tomar inicialmente $m = 1$. Desta forma temos $a_1\sigma_1(w) = 0$ para todo $w \in K$, logo $a_1 = 0$, o que contradiz a hipótese. Assim o teorema é válido para $m = 1$. Vamos assumir que $m > 1$. Como os automorfismos são distintos, existe um elemento $c \in K$ tal que $\sigma_1(c) \neq \sigma_m(c)$. Como $cw \in K$ para todo $w \in K$, a igualdade

$$\begin{aligned} a_1\sigma_1(cw) + a_2\sigma_2(cw) + \cdots + a_n\sigma_n(cw) &= 0 \\ a_1\sigma_1(c)\sigma_1(w) + a_2\sigma_2(c)\sigma_2(w) + \cdots + a_n\sigma_n(c)\sigma_n(w) &= 0 \end{aligned} \tag{6.5}$$

também deve ser válida para todo $w \in K$. Entretanto, se partirmos de

$$a_1\sigma_1(w) + a_2\sigma_2(w) + \cdots + a_m\sigma_m(w) = 0 \tag{6.6}$$

e multiplicarmos (6.6) por $\sigma_1(c)$, chegamos em

$$a_1\sigma_1(c)\sigma_1(w) + a_2\sigma_1(c)\sigma_2(w) + \cdots + a_m\sigma_1(c)\sigma_m(w) = 0. \tag{6.7}$$

Agora subtraímos a equação (6.5) da equação (6.7),

$$\begin{aligned} a_2\sigma_2(c)\sigma_2(w) - a_2\sigma_1(c)\sigma_2(w) + \cdots + a_m\sigma_m(c)\sigma_m(w) - a_m\sigma_1(c)\sigma_m(w) &= 0 \\ a_2\left(\sigma_2(c) - \sigma_1(c)\right)\sigma_2(w) + \cdots + a_m\left(\sigma_m(c) - \sigma_1(c)\right)\sigma_m(w) &= 0. \end{aligned}$$

Se colocarmos $b_i = a_i\left(\sigma_i(c) - \sigma_1(c)\right)$ para $i = 2, \ldots, m$, então os b_i estão em K, $b_m = a_m\left(\sigma_m(c) - \sigma_1(c)\right) \neq 0$, pois $a_m \neq 0$ e $\sigma_m(c) - \sigma_1(c) \neq 0$, mas $b_2\sigma_2(w) + \cdots + b_m\sigma_m(w) = 0$ para todo $w \in K$. O que mais uma vez contradiz a hipótese e portanto o teorema está demonstrado. □

Definição 6.29. Se G é um grupo de automorfismos de K, então o *corpo fixo* $\mathcal{G}$ *de* G é o conjunto de todos os elementos $a \in K$ tais que $\sigma(a) = a$ para todo $\sigma \in G$.

Lema 6.30. *O corpo fixo* $\mathcal{G}$ *de* G *é um subcorpo de* K.

Demonstração. Sejam $a, b \in \mathcal{G}$, onde $\mathcal{G}$ é um corpo fixo de G grupo de automorfismos. Assim para todo $\sigma \in G$, temos $\sigma(a) = a$ e $\sigma(b) = b$. Portanto, $\sigma(a \pm b) = \sigma(a) \pm \sigma(b) = a \pm b$ e $\sigma(ab) = \sigma(a)\sigma(b) = ab$. Logo $a \pm b$ e ab estão no corpo fixo $\mathcal{G}$ de G. Se $b \neq 0$, então $\sigma\left(b^{-1}\right) = \sigma\left(b\right)^{-1} = b^{-1}$, donde b^{-1} também está no corpo fixo $\mathcal{G}$ de G. □

Definição 6.31. Sejam K um corpo e F um subcorpo de K. Então o *grupo dos automorfismos de* K *relativos a* F, indicado por $G(K, F)$, é o conjunto de todos os automorfismos de K que deixam fixo todo elemento de F, isto é, o automorfismo σ de K está em $G(K, F)$ se, e somente se, $\sigma(\alpha) = \alpha$ para todo $\alpha \in F$.

Lema 6.32. $G(K, F)$ *é um subgrupo do grupo de todos os automorfismos de* K.

Demonstração. Pelo Lema 4.10, precisamos mostrar que para todo $\sigma, \tau \in G(K, F)$, temos $\sigma \circ \tau \in G(K, F)$ e que $\sigma^{-1} \in G(K, F)$. Tomando $\alpha \in F$, temos $(\sigma \circ \tau)(\alpha) = \tau\left(\sigma(\alpha)\right) = \tau(\alpha) = \alpha$ e portanto $\sigma \circ \tau \in G(K, F)$. E claramente, $\sigma^{-1} \in G(K, F)$. Pois $\left(\sigma^{-1} \circ \sigma\right)(\alpha) = \iota(\alpha) = \alpha$, onde ι é a função identidade. □

Vejamos alguns exemplos dos grupos que definimos a pouco.

Exemplo 6.7. Considere o grupo $G(\mathbb{C}, \mathbb{R})$ de automorfismos dos números complexos sobre os números reais. Este grupo é formado por todos os automorfismos que deixam os números reais fixos. Assim, se tomarmos $\sigma \in G(\mathbb{C}, \mathbb{R})$ e $a + bi \in \mathbb{C}$, com $a, b \in \mathbb{R}$, temos

$$\sigma(a + bi) = \sigma(a) + \sigma(b)\sigma(i) = a + b\sigma(i).$$

Como $i^2 = -1$, temos

$$\sigma(i) \circ \sigma(i) = \sigma(i^2) = \sigma(-1) = -1,$$

assim $\sigma_1(i) = i$ ou $\sigma_2(i) = -i$. Desta forma, $G(\mathbb{C}, \mathbb{R}) = \{\sigma_1, \sigma_2\}$. Como $\sigma_1(a+bi) = \sigma_1(a) + \sigma_1(b)\sigma_1(i) = a + bi$, temos que σ_1 é o elemento identidade do grupo. Observamos também, que $\sigma_2(a+bi) = \sigma_2(a) + \sigma_2(b)\sigma_2(i) = a - bi$, e consequentemente σ_2 é a conjugação complexa e portanto $G(\mathbb{C}, \mathbb{R})$ é um grupo de ordem 2.

Exemplo 6.8. Sejam $G\left(\mathbb{Q}(\sqrt[3]{2}), \mathbb{Q}\right)$ e $\sigma \in G\left(\mathbb{Q}(\sqrt[3]{2}), \mathbb{Q}\right)$. Assim σ deixa todo elemento $\alpha \in \mathbb{Q}$ fixo. Lembremo-nos que $\mathbb{Q}(\sqrt[3]{2})$ é uma extensão de $\mathbb{Q}$ e que $\left[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}\right] = 3$, pois $\mathcal{B} = \{1, \sqrt[3]{2}, \sqrt[3]{4}\}$ é uma base para $\mathbb{Q}(\sqrt[3]{2})$. Assim podemos escrever um elemento arbitrário de $\mathbb{Q}(\sqrt[3]{2})$ como sendo $\alpha_1 + \alpha_2\sqrt[3]{2} + \alpha_3\sqrt[3]{4}$, com $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{Q}$. Entretanto, temos que

$$\sigma\left(\sqrt[3]{2}\right) \circ \sigma\left(\sqrt[3]{4}\right) = \sigma\left(\sqrt[3]{2}\sqrt[3]{4}\right) = \sigma(2) = 2,$$

o que implica que $\sigma\left(\sqrt[3]{2}\right) = \sqrt[3]{2}$ e $\sigma\left(\sqrt[3]{4}\right) = \sqrt[3]{4}$. Desta forma, para todos os elementos $\alpha_1 + \alpha_2\sqrt[3]{2} + \alpha_3\sqrt[3]{4} \in \mathbb{Q}(\sqrt[3]{2})$, com $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{Q}$, temos

$$\sigma\left(\alpha_1 + \alpha_2\sqrt[3]{2} + \alpha_3\sqrt[3]{4}\right) = \sigma(\alpha_1) + \sigma(\alpha_2)\sigma(\sqrt[2]{3}) + \sigma(\alpha_3)\sigma(\sqrt[2]{4}) = \alpha_1 + \alpha_2\sqrt[2]{3} + \alpha_3\sqrt[2]{4}.$$

Vemos assim que $G\left(\mathbb{Q}(\sqrt[3]{2}), \mathbb{Q}\right) = \{\sigma\}$, onde σ é o elemento identidade do grupo. Neste caso, o *corpo fixo* de $G\left(\mathbb{Q}(\sqrt[3]{2}), \mathbb{Q}\right)$ *não é* $\mathbb{Q}$ *mas é, de fato, maior, sendo todo o* $\mathbb{Q}(\sqrt[3]{2})$.

O próximo teorema nos mostrará que $G(K, F)$ tem um limite importante sobre o seu tamanho.

Teorema 6.33. *Se K é uma extensão finita de F, então $G(K, F)$ é um grupo finito e sua ordem $o(G(K, F))$ satisfaz $o(G(K, F)) \leq [K : F]$.*

Demonstração. Considere $[K : F] = n$ e que $\mathcal{B} = \{w_1, \ldots, w_n\}$ seja uma base de K sobre F. Suponhamos que seja possível encontrar $n+1$ automorfismos distintos $\sigma_1, \sigma_2, \ldots, \sigma_{n+1}$ em $G(K, F)$. Assim podemos escrever:

$$\begin{cases} \sigma_1(w_1)x_1 + \sigma_2(w_1)x_2 + \cdots + \sigma_{n+1}(w_1)x_{n+1} = 0 \\ \vdots \\ \sigma_1(w_i)x_1 + \sigma_2(w_i)x_2 + \cdots + \sigma_{n+1}(w_i)x_{n+1} = 0 \\ \vdots \\ \sigma_1(w_n)x_1 + \sigma_2(w_n)x_2 + \cdots + \sigma_{n+1}(w_n)x_{n+1} = 0. \end{cases}$$

Assim o sistema possui n equações com $n+1$ incógnitas e portanto existe uma solução não trivial $x_1 = a_1, \ldots, x_{n+1} = a_{n+1}$ (não todos nulos) em K. Assim

$$a_1\sigma_1(w_i) + a_2\sigma_2(w_i) + \cdots a_{n+1}\sigma_{n+1}(w_i) = 0, \tag{6.8}$$

para $i = 1, 2, \ldots, n$.

Como todo elemento em F é deixado fixo em relação a cada σ_i e para um $t \in K$ qualquer, temos

$$t = \alpha_1 w_1 + \cdots + \alpha_n w_n,$$

com $\alpha_1, \ldots, \alpha_n$ em F. Então pelo sistema de equações (6.8) obtemos

$$a_1\sigma_1(t) + \cdots a_{n+1}\sigma_{n+1}(t) = 0,$$

para todo $t \in K$. Mas isto contradiz o Teorema 6.28, o que demonstra o resultado. □

De modo semelhante ao que fizemos no Capítulo 4, Subseção 4.3.4 onde definimos o anel dos polinômios nas variáveis $x_1, \ldots, x_n$ sobre F, definimos o corpo $F(x_1, \ldots, x_n)$ das funções racionais em $x_1, \ldots, x_n$ sobre F como o anel de todas as frações de tais polinômios, onde o denominador não é um polinômio identicamente nulo. Como $F[x]$ é um anel de integridade, podemos construir tal corpo. Este corpo consiste simplesmente de todas as frações de polinômios e é denominado *corpo de funções racionais em x sobre F*.

Considere o grupo simétrico S_n de grau n. Considere uma função $\sigma \in S_n$ e i um inteiro com $1 \leq i \leq n$, onde $\sigma(i)$ é a imagem de i por meio de σ. Faremos S_n operar sobre $F(x_1, \ldots, x_n)$. Assim:

Definição 6.34. Definimos para $\sigma \in S_n$ e $r(x_1, \ldots, x_n) \in F(x_1, \ldots, x_n)$ a função que leva $r(x_1, \ldots, x_n)$ em $r(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$. Estas funções definem automorfismos de $F(x_1, \ldots, x_n)$ que também chamarei de σ. Além do mais, o corpo fixo de $F(x_1, \ldots, x_n)$ consiste de todas as funções racionais $r(x_1, \ldots, x_n)$, tais que $r(x_1, \ldots, x_n) = r(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ para todo $\sigma \in S_n$. Um subcorpo de $F(x_1, \ldots, x_n)$ sobre um corpo fixo de S_n é denominado *corpo das funções racionais simétricas* e indicaremos por S.

Podemos explicitar em S certas funções particularmente simples, que são construídas a partir de $x_1, \ldots, x_n$, conhecidas como *funções simétricas elementares em* $x_1, \ldots, x_n$. Vejamos alguns exemplos de como podemos proceder para construir tais funções.

Exemplo 6.9. Na Seção 1.2 encontramos as raízes x_1 e x_2 da equação $x^2 - sx + p = 0$ onde s é a soma destas raízes, ou seja, $s = x_1 + x_2$ e $p = x_1x_2$ o seu produto. Tomemos o polinômio $q(t) = t^2 - st + p$, com as raízes em $F(x_1, x_2)$, onde x_1 e x_2 são as raízes

do polinômio $q(t)$. Queremos expressar $q(t)$ com coeficientes em função das raízes do polinômio. O que neste caso é bem simples, pois s e p já estão em função das raízes de $q(x)$. Assim, $a_1 = s$ e $a_2 = p$ são chamadas funções simétricas elementares. Deste modo, podemos escrever $F(x_1, x_2)$ como sendo o corpo das raízes do polinômio $t^2 - a_1 t + a_2$ em $F(a_1, a_2)$.

Exemplo 6.10. Para $n = 3$, tomando x_1, x_2 e x_3 as raízes de um polinômio $p(t)$. Observamos que, podemos escrever $p(t)$ como um polinômio sobre $F(a_1, a_2, a_3)$, onde $a_1 = x_1 + x_2 + x_3$, $a_2 = x_1x_2 + x_1x_3 + x_2x_3$ e $a_3 = x_1x_2x_3$ são as funções elementares simétricas. Note que:

$$\begin{aligned}(t - x_1)(t - x_2)(t - x_3) &= (t^2 - (x_1 + x_2)t + x_1x_2)(t - x_3)\\ &= t^3 - (x_1 + x_2)t^2 + x_1x_2t - x_3t^2 + (x_1x_3 + x_2x_3)t - x_1x_2x_3\\ &= t^3 - (x_1 + x_2 + x_3)t^2 + (x_1x_2 + x_1x_3 + x_2x_3)t - x_1x_2x_3.\end{aligned}$$

Logo, $p(t) = t^3 - a_1t^2 + a_2t - a_3$ é um polinômio sobre $F(a_1, a_2, a_3)$ e suas raízes x_1, x_2 e x_3 estão em $F(x_1, x_2, x_3)$.

De modo geral, podemos fatorar $p(t) = t^n - a_1t^{n-1} + a_2^{n-2} - \cdots + (-1)^n a_n$, que têm coeficientes em $F(a_1, \ldots, a_n)$, como $p(t) = (t - x_1)(t - x_2)\cdots(t - x_n)$ sobre $F(x_1, \ldots, x_n)$, com $x_1, \ldots, x_n$ raízes de $p(t)$, onde:

$$\begin{aligned}a_1 &= x_1 + x_2 + \cdots + x_n = \sum_{i=1}^{n} x_i\\ a_2 &= \sum_{i<j} x_ix_j\\ a_3 &= \sum_{i<j<k} x_ix_jx_k\\ &\vdots\\ a_n &= x_1x_2\cdots x_n.\end{aligned}$$

Agora temos $a_1, \ldots, a_n$ em S e com isso obtemos o corpo $F(a_1, \ldots, a_n) \subset S$ pela adjunção de $a_1, \ldots, a_n$ a F. Com isto, um dos nossos objetivos principais no teorema que demonstraremos a seguir é mostrar que $S = F(a_1, \ldots, a_n)$.

Teorema 6.35. *Sejam F um corpo e $F(x_1, \ldots, x_n)$ o corpo das funções racionais em $x_1, \ldots, x_n$ sobre F. Suponhamos que S seja o corpo das funções racionais simétricas, então*

1. $[F(x_1,\ldots,x_n):S]=n!$
2. $G\left(F(x_1,\ldots,x_n),S\right)=S_n$, *o grupo simétrico de grau n.*
3. *Se* $a_1,\ldots,a_n$ *são as funções simétricas elementares, então* $S=F(a_1,a_2,\ldots,a_n)$, *em* $x_1,\ldots,x_n$.
4. $F(x_1,\ldots,x_n)$ *é o corpo de raízes sobre* $F(a_1,a_2,\ldots,a_n)=S$ *do polinômio* $t^n-a_1t^{n-1}+a_2t^{n-2}+\cdots+(-1)^na_n$.

Demonstração. Apesar de separar em itens, eles se misturam na demonstração. Deste modo, veremos abaixo que para mostrar os itens 1 e 3, acabamos mostrando 2 e 4. Assim, temos:

1. Como o grupo S_n é um grupo de automorfismos de $F(x_1,\ldots,x_n)$ que deixa S fixo, $S_n\subset G\left(F(x_1,\ldots,x_n),S\right)$. Assim, pelo Teorema 6.33, $[F(x_1,\ldots,x_n):S]\geq o\left(G\left(F(x_1,\ldots,x_n),S\right)\right)\geq o(S_n)=n!$. Por outro lado, a prova do item 3 nos garante que $[F(x_1,\ldots,x_n):F(a_1,\ldots,a_n)]\leq n!$, pois $F(a_1,\ldots,a_n)$ é um subcorpo de S, e assim $n!\geq[F(x_1,\ldots,x_n):F(a_1,\ldots,a_n)]=[F(x_1,\ldots,x_n):S]\,[S:F(a_1,\ldots,a_n)]\geq n!$ e portanto, temos $[F(x_1,\ldots,x_n):S]=n!$, pois $[S:F(a_1,\ldots,a_n)]=1$ e $S=F(a_1,\ldots,a_n)$.
2. O exposto no item anterior, nos garante que $G\left(F(x_1,\ldots,x_n),S\right)=S_n$, mas para isto devemos provar que $[F(x_1,\ldots,x_n):F(a_1,\ldots,a_n)]\leq n!$ e com isso demostramos 3.
3. Para mostrarmos que $[F(x_1,\ldots,x_n):F(a_1,\ldots,a_n)]\leq n!$, notemos que o polinômio $p(t)=t^n-a_1t^{n-1}+a_2^{n-2}-\cdots+(-1)^na_n$, que tem coeficiente em $F(a_1,\ldots,a_n)$, fatora-se sobre $F(x_1,\ldots,x_n)$ como $p(t)=(t-x_1)(t-x_2)\cdots(t-x_n)$. Assim, $p(t)$, de grau n sobre $F(a_1,\ldots,a_n)$, fatora-se como um produto de fatores lineares sobre $F(x_1,\ldots,x_n)$. Não se pode fatorar $p(t)$ sobre um subcorpo próprio de $F(x_1,\ldots,x_n)$ que contém $F(a_1,\ldots,a_n)$, pois este subcorpo teria então de conter F e cada uma das raízes, $x_1,\ldots,x_n$ em $p(t)$, portanto, este subcorpo coincidiria com $F(x_1,\ldots,x_n)$. Assim, vemos que $F(x_1,\ldots,x_n)$ é o corpo de raízes do polinômio $p(t)=t^n-a_1t^{n-1}+a_2^{n-2}-\cdots+(-1)^na_n$ sobre $F(a_1,\ldots,a_n)$. Como $p(t)$ é de grau n, pelo Corolário 6.16, obtemos $[F(x_1,\ldots,x_n):F(a_1,\ldots,a_n)]\leq n!$
4. Este item foi demonstrado no anterior.

E portanto o teorema está demonstrado. □

Grupos de automorfismos de corpos e de corpos fixos com relação a tais grupos, podem ter F menor que todo o corpo fixo $G(K, F)$. Seguramente F está sempre contido neste corpo, mas não o esgota necessariamente. Seria interessante que uma extensão K de F em que F seja precisamente o corpo fixo de $G(K, F)$, assim esta condição é uma autêntica limitação sobre o tipo de extensão de F que estamos considerando.

Isto nos leva a definir *extensão normal*, como:

Definição 6.36. Dizemos que K é uma *extensão normal* de F, se K é uma extensão finita de F tal que F seja o corpo fixo de $G(K, F)$.

Em outras palavras, se K é uma extensão normal de F, então todo elemento em K que está fora de F é movido por algum elemento em $G(K, F)$. Podemos ver que no Exemplo 6.7 temos uma extensão normal e no Exemplo 6.8 não.

Agora podemos refinar nossa definição de grupo de Galois, o que a propósito nos permitirá seguir em nossa estratégia em relacionar o grupo de Galois com grupos solúveis.

Definição 6.37. Sejam $f(x)$ um polinômio em $F[x]$ e K seu corpo de raízes sobre F. O *grupo de Galois* de $f(x)$ é o grupo $G(K, F)$ de todos os automorfismos de K que deixam fixo todo elemento de F.

Podemos relacionar a resolução por radicais de $p(x)$ com a solubilidade, como um grupo, do grupo de Galois de $p(x)$.

Lema 6.38. *Suponhamos que todas as raízes enésimas da unidade (para um certo n particular) estejam em F e suponhamos que $a \neq 0$ esteja em F. Sejam $x^n - a \in F[x]$ e K seu corpo de raízes sobre F. Então:*

1. *$K = F(u)$, onde u é qualquer raiz de $x^n - a$.*
2. *O grupo de Galois de $x^n - a$ sobre F é abeliano.*

Chegamos a um teorema chave para que possamos estabelecer as condições para a resolução por radicais das raízes de um polinômio. Este resultado faz uma correspondência biunívoca entre os subcorpos dos corpos de raízes de $f(x)$ e os subgrupos de seu

grupo de Galois. E ainda mais, ele nos dá um critério para verificar se um subcorpo de uma extensão normal é também uma extensão normal de F. Estamos falando do *Teorema Fundamental da Teoria de Galois*. O Lema 6.38 e o Teorema 6.39 (Fundamental da Teoria de Galois) estão demonstrados em [6].

Teorema 6.39. *Fundamental da Teoria de Galois*
Sejam $f(x)$ um polinômio em $F[x]$, K seu corpo de raízes sobre F e $G(K,F)$ seu grupo de Galois. Para todo subgrupo T de K que contém F seja $G(K,T) = \{\sigma \in G(K,F);\ \sigma(t) = t \text{ para todo } t \in T\}$ e para todo subgrupo H de $G(K,F)$ seja $K_H = \{x \in K;\ \sigma(x) = x \text{ para todo } \sigma \in H\}$. Então a associação de T com $G(K,T)$ estabelece uma correspondência bijetiva do conjunto dos subcorpos de K que contêm F sobre o conjunto dos subgrupos de $G(K,F)$ tal que

1. $T = K_{G(K,T)}$.
2. $H = G(K, K_H)$.
3. $[K : T] = o\left(G(K,T)\right)$, $[T : F] =$ *índice de* $G(K,T)$ *em* $G(K,F)$.
4. *T é uma extensão normal de F se, e somente se, $G(K,T)$ é um subgrupo normal de $G(K,F)$.*
5. *Quando T é uma extensão normal de F, $G(T,F)$ é isomorfo a $G(K,F)/G(K,T)$.*

O Teorema 6.39 Fundamental da Teoria de Galois nos traz a principal condição para relacionar o grupo de Galois e os grupos solúveis.

Teorema 6.40. *Se $p(x) \in F[x]$ é resolúvel por radicais sobre F, então o grupo de Galois sobre F de $p(x)$ é um grupo solúvel.*

Demonstração. Seja K o corpo de raízes de $p(x)$ sobre F, o grupo de Galois de $p(x)$ sobre F é $G(K,F)$. Considere $F_K \supset K$ uma extensão normal de F. Assim, se $\omega_1, \omega_2, \ldots, \omega_k$ são as raízes de $p(x)$. Pelo Lema 6.38 podemos escrever $F_1 = F(\omega_1)$, onde F_1 é uma extensão normal de F que contém a raiz ω_1 e $\omega_1^{r_1} \in F$. Também podemos construir $F_2 = F_1(\omega_2)$ uma extensão normal não somente de F_1, mas também de F. Seguindo este raciocínio chegamos à cadeia

$$F \subset F_1 = F(\omega_1) \subset F_2 = F_1(\omega_2) \subset \cdots \subset F_k = F_{k-1}(\omega_k),$$

onde $\omega_1^{r_1} \in F$, $\omega_2^{r_2} \in F_1$, $\ldots$, $\omega_k^{r_k} \in F_{k-1}$.

Como F_K é uma extensão normal de F, F_K também é uma extensão normal de qualquer corpo intermediário, logo F_K é uma extensão normal de cada F_i.

Pelo Lema 6.38 temos $G(F_K, F)$ é abeliano e portanto F_i é uma extensão normal de F_{i-1}. Pelo Teorema 6.39 *Fundamental da Teoria de Galois*, como F_i é uma extensão normal sobre F_K, o subgrupo $G(F_i, F_k)$ é um subgrupo normal de $G(F_i, F_{i-1})$ que é isomorfo a $G(F_K, F_{i-1})/G(F_K, F_i)$. Assim podemos construir a cadeia

$$G(F_K, F) \supset G(F_K, F_1) \supset G(F_K, F_2) \supset \cdots \supset G(F_K, F_{k-1}) \supset \{e\}. \tag{6.9}$$

Pelo Lema 6.38 temos que $G(F_i, F_{i-1})$ é abeliano, deste modo todo o grupo quociente $G(F_K, F_{i-1})/G(F_K, F_i)$ da cadeia (6.9) é abeliano. E consequentemente o grupo $G(F_K, F)$ é solúvel.

Sabendo que o corpo de raízes K é uma extensão normal de F e que $F_K \supset K$, temos pelo Teorema 6.39 *Fundamental da Teoria de Galois* que o $G(F_K, K)$ é um subgrupo normal de $G(F_K, F)$. Também sabemos que $G(F_K, F)$ é isomorfo a $G(F_K, F)/G(F_K, K)$. Assim $G(K, F)$ é uma imagem homomorfa de $G(F_K, F)$ que é um grupo solúvel e pelo Corolário 6.22, se $G(F_K, F)$ é um grupo solúvel, então $G(K, F)$, que é o grupo de Galois de $p(x)$ sobre F, também é um grupo solúvel. □

Finalmente apresentamos o Teorema 6.41 atribuído a Abel [2] que conclui a impossibilidade de determinar as raízes de um polinômio $p(x) = x^n + a_1 x^{n-1} + \cdots + a_n$ de grau $n \geq 5$.

Teorema 6.41. *O polinômio geral de grau* 5 *não é resolúvel por radicais.*

Demonstração. Pelo Teorema 6.35, $F(a_1, \ldots, a_n)$ é o corpo das funções racionais nas n variáveis $a_1, \ldots, a_n$. Então o grupo de Galois do polinômio $p(t) = t^n + a_1 t^{n-1} + \cdots + a_n$ sobre $F(a_1, \ldots, a_n)$ é S_n, ou seja, o grupo simétrico de grau n. Entretanto, pelo Teorema 6.24, S_n não é um grupo solúvel para $n \geq 5$. Desta forma, usando a contra-positiva do Teorema 6.40, temos que por S_n não ser solúvel, $p(x)$ não é resolúvel por radicais sobre $F(a_1, \ldots, a_n)$ para $n \geq 5$. □

2 Niels Henrik Abel foi um matemático norueguês, que nasceu em *Nedstrand*, 5 de agosto de 1802 e faleceu em *Froland*, 6 de abril de 1829.

7

MÉTODO DE NEWTON

Nos dedicamos nos primeiros capítulos a mostrar as soluções das equações quadráticas, cúbicas e quárticas. E bem mais que um capítulo, foi necessário para mostrar que uma equação de quinto grau ou superior não têm uma fórmula para sua solução. Entretanto, isto não significa que não possamos encontrar raízes para tais equações com uma precisão tão boa quanto se queira. É nisto que nos dedicaremos neste capítulo. Apresentaremos o método de *Newton-Raphson*,[1] que não só resolve equações polinomiais, mas também equações transcendentais, tais como $\cos x = x$.

Entretanto, o nosso objetivo principal é mostrar que é possível determinar uma boa aproximação da raiz real do polinômio $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$, com $n \geq 5$. Para determinar tal raiz, se considerarmos um intervalo $[a, b]$ do polinômio p, onde saibamos que $p(a)$ e $p(b)$ possuem sinais contrários é certo que uma raiz r deste polinômio está entre a e b. O procedimento para se obter uma aproximação para a raiz r é obter uma sequência $x_1, x_2, \ldots, x_n, \ldots$ de números que se aproximem de r. Uma vez determinado x_1 que está entre a e b, verificamos se $p(x_1)$ é maior ou menor do que zero. É óbvio que ser for igual a zero $x_1 = r$ e o problema termina. Suponhamos que $p(x_1)$ e $p(b)$ sejam ambos maior do que zero, assim $a < r < x_1 < b$. Agora podemos tomar o intervalo $[a, x_1]$ e proceder de modo análogo.

Contudo, os valores não são tomados arbitrariamente. Este processo iterativo leva em consideração uma reta tangente à função num certo ponto $P(x_1, y_1)$, onde o x_1 é a

1 Isaac Newton foi um cientista inglês, mais reconhecido como físico e matemático, embora tenha sido também astrônomo, alquimista, filósofo natural e teólogo. Nasceu em *Woolsthorpe-by-Colsterworth*, no dia 25 de dezembro de 1642 e faleceu em *Kensington*, no dia 20 de março de 1726.
Joseph Raphson foi um matemático inglês, pouco se sabe sobre sua vida, o ano de seu nascimento e morte foram aproximados pelo historiador matemático Florian Cajori.(c.1648-c.1715)

primeira aproximação para a raiz. Assim a reta tangente será um ponto chave para a execução deste método.

7.1 NOÇÕES DE CÁLCULO

Para aplicarmos este método usaremos algumas noções de cálculo. Então começaremos com algumas definições retiradas de [13]. Iniciaremos com a noção de *limite* que está relacionado a determinação de uma reta tangente a uma curva dada.

7.1.1 *Limite*

Definição 7.1. Dizemos que L é o *limite* de uma função f sobre algum intervalo aberto que contém o número a, exceto possivelmente no próprio a, e denotaremos o limite L de $f(x)$ quando x tende a a por

$$\lim_{x \to a} f(x) = L,$$

se para todo número $\varepsilon > 0$ há um número correspondente $\delta > 0$ tal que

$$| f(x) - L | < \varepsilon \qquad \text{sempre que} \qquad 0 < | x - a | < \delta.$$

Podemos dizer, uma vez que $| x - a |$ é a distância de x a a e $| f(x) - L |$ é a distância de $f(x)$ a L, e como ε pode ser arbitrariamente pequeno, que $\lim_{x \to a} f(x) = L$ significa que a distância entre $f(x)$ e L pode ser arbitrariamente pequena tornando-se a distância de x a a suficientemente pequena, mas não 0.

7.1.2 *Tangentes*

Como já foi dito um ponto fundamental para este método são as retas tangentes a uma curva. Para isto, tomemos uma curva C de equação $y = f(x)$. Para calcular a tangente de C em um ponto $P\,(a, f(a))$, consideramos um ponto $Q\,(x, f(x))$, onde $x \neq a$, e calculamos a inclinação da reta secante PQ:

$$m_{PQ} = \frac{f(x) - f(a)}{x - a}.$$

Então fazemos Q aproximar-se de P ao longo da curva C ao obrigar x tender a a (Figura 14).

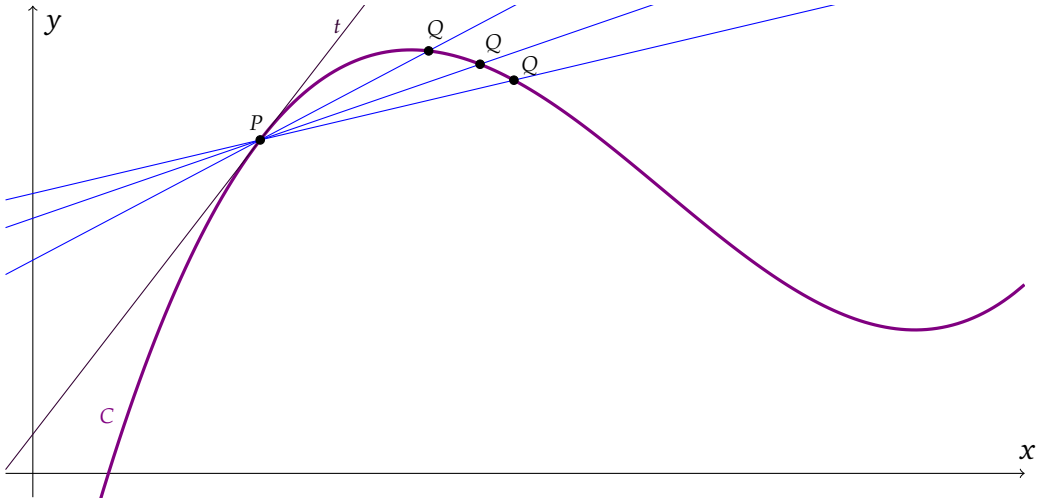

Figura 14: Q se aproximando de P.

Se m_{PQ} tender a um número m, então definimos a *tangente* t como sendo a reta que passa por P e tem inclinação m. A Figura 15 ilustra que a reta tangente é a posição limite da reta secante PQ quando Q tende a P.

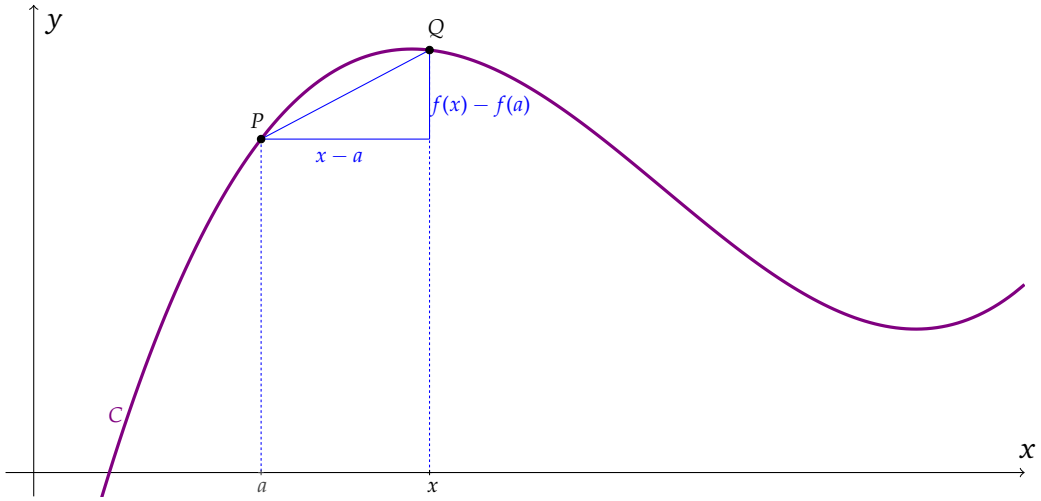

Figura 15: PQ quando Q tende a P.

Assim definimos

Definição 7.2. A *reta tangente* a uma curva $y = f(x)$ em um ponto $P\,(a, f(a))$ é a reta por P que tem inclinação

$$m = \lim_{x \to a} \frac{f(x) - f(a)}{x - a},$$

desde que esse limite exista.

Podemos reescrever a expressão para a inclinação da reta tangente (Figura 16). Tome $h = x - a$, assim $x = a + h$. Logo a inclinação da reta PQ é

$$m_{PQ} = \frac{f(a+h) - f(a)}{h}.$$

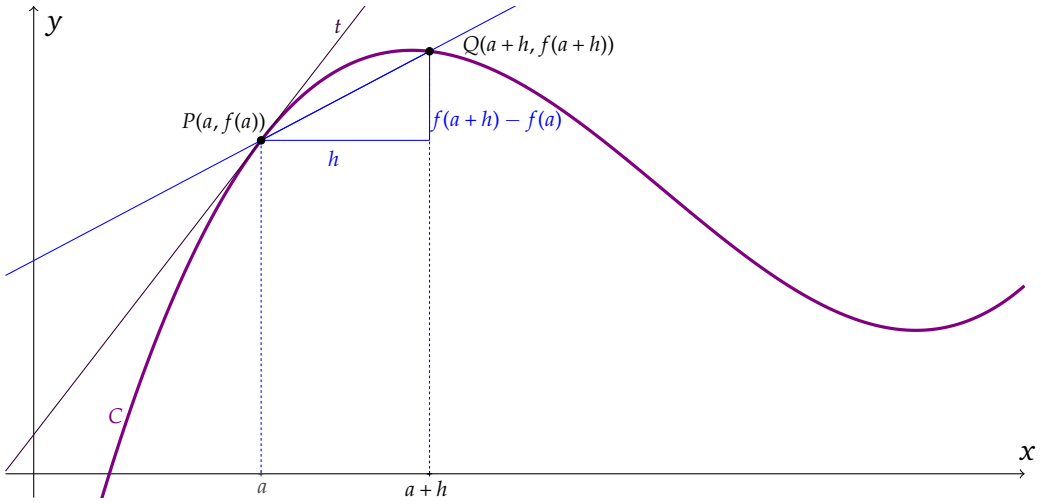

Figura 16: Neste caso $h > 0$ e Q está à direita de P. No caso de $h < 0$, o ponto Q estará à esquerda de P.

Deste modo, se x tende a a, h tende a 0 (pois $h = x - a$). Portanto, a expressão para a inclinação da reta tangente da Definição 7.2 pode ser escrita como:

$$m = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

7.1.3 *Derivadas*

A definição de *derivada*, está estreitamente relacionada com a noção de reta tangente que definimos a pouco.

Definição 7.3. A *derivada* de uma função f em um número a, denotada por $f'(a)$, é

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h},$$

se o limite existe.

Assim como fizemos antes, podemos aqui também tomar $x = a + h$, assim $h = x - a$. Desta maneira h tende a 0 se, e somente se, x aproximar-se de a. Consequentemente, uma maneira equivalente de enunciar a Definição 7.3 da derivada, como vimos na determinação das retas tangentes, é

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}.$$

Percebemos deste modo, que a reta tangente a $y = f(x)$ em $(a, f(a))$ é a reta que passa em $(a, f(a))$, cuja inclinação é igual a $f'(a)$, ou seja, a derivada de f em a (Figura 17).

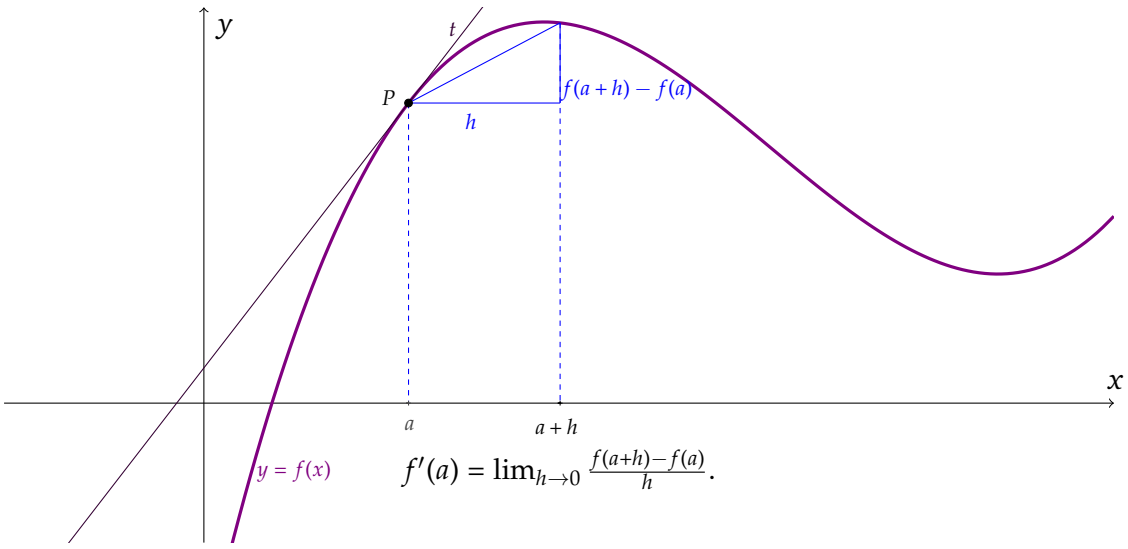

Figura 17: $f'(a)$ tem a mesma inclinação da tangente em P que é a mesma inclinação da curva em P. Quando dizemos a inclinação da curva, estamos considerando que estamos tomando um intervalo muito pequeno da curva, a tal ponto que se ampliássemos este pedaço ele seria semelhante a uma reta. É essa inclinação da curva de que nos referimos.

7.2 O MÉTODO DE NEWTON-RAPHSON

Vamos aplicar o método para encontrar uma aproximação da raiz r de uma função f. Para isto, vamos exibir, conforme Figura 18, o gráfico de f e uma primeira aproximação x_1 que pode ser obtida através de uma conjectura, ou de um esboço do gráfico de f.

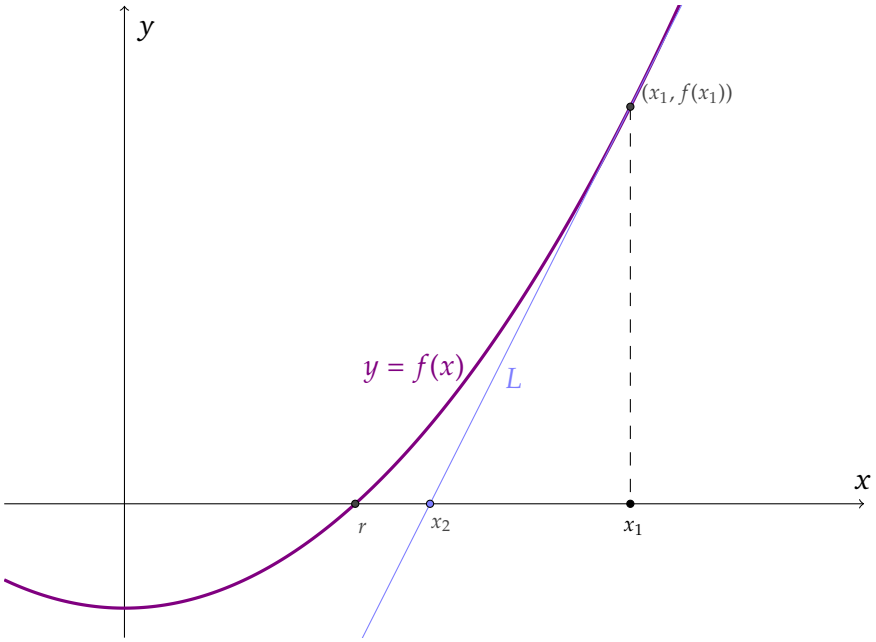

Figura 18: Gráfico da função f com sua raiz r. Atribuímos x_1 uma aproximação inicial e com ela chegamos a um valor x_2 mais próximo de r através da reta tangente L.

O método de Newton-Raphson é um processo iterativo que consiste em tomarmos sucessivamente retas tangentes a determinados pontos [13]. Faremos isto do seguinte modo, vamos primeiramente tomar a reta tangente L à curva $y = f(x)$ no ponto $(x_1, f(x_1))$. A interseção desta reta L com o eixo x se aproxima mais de r conforme observamos na Figura 18 e por isso x_2 é uma aproximação para r melhor do que x_1.

Para calcular x_2, o escrevemos em termos de x_1, usamos o fato de que a inclinação de L é $f'(x_1)$. Assim, sua equação é:

$$y - f(x) = f'(x_1)(x - x_1).$$

Uma vez que L intersecta o eixo x em x_2, fazendo $y = 0$, obtemos

$$0 - f(x_1) = f'(x_1)(x_2 - x_1).$$

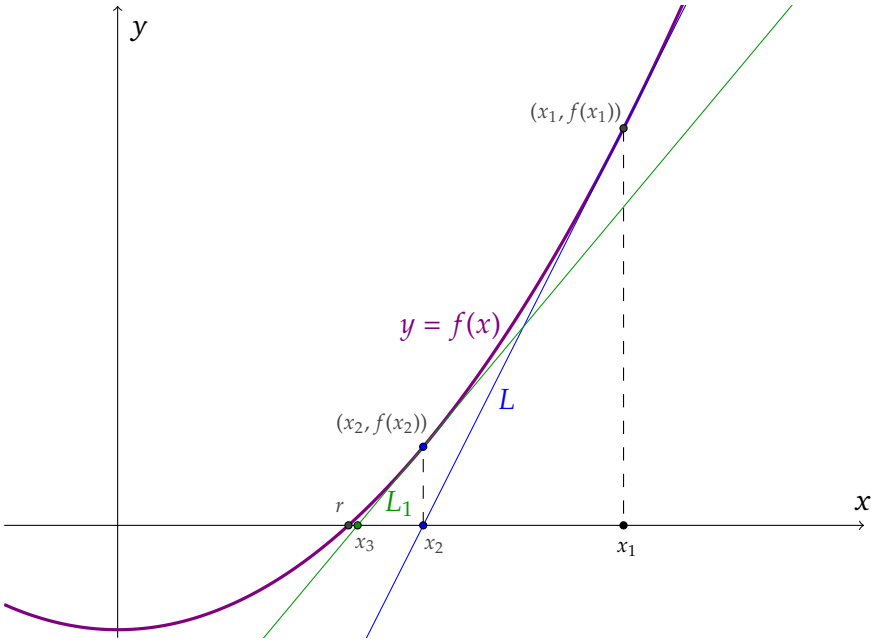

Figura 19: Traçamos a reta tangente L_1 pelo ponto $(x_2, f(x_2))$ para encontrarmos a terceira aproximação para raiz r da função $f(x)$.

Considerando $f'(x_1) \neq 0$, podemos resolver essa equação para x_2:

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}.$$

Agora procederemos de modo análogo para determinar uma aproximação x_3 ainda melhor para a raiz r. Observe a Figura 19.

Repetiremos o mesmo procedimento que fizemos anteriormente, assim temos:

$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)}.$$

Como já dissemos, este é um método iterativo. Assim, repetindo esse processo obteremos uma sequência de aproximações $x_1, x_2, x_3, x_4, \ldots$. Na Figura 20 é mostrada a aproximação x_4. Em geral, se x_n é a enésima aproximação e $f'(x_n) \neq 0$, então a aproximação seguinte é dada por:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{7.1}$$

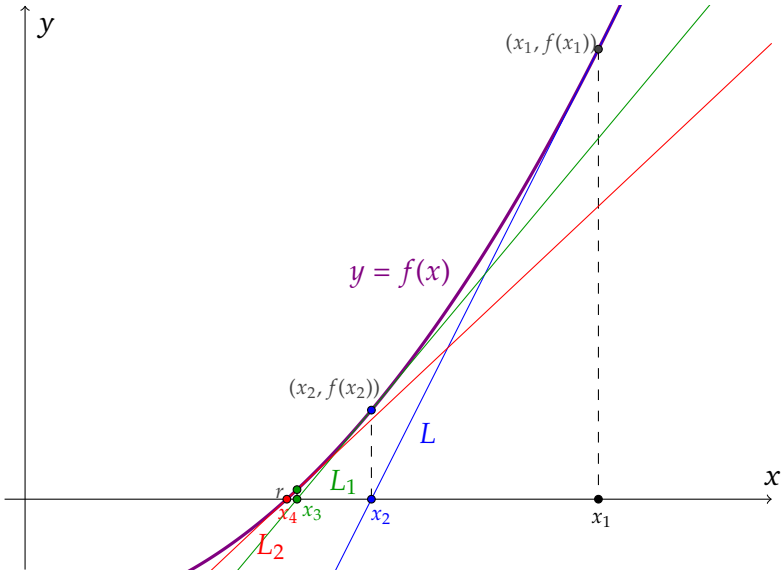

Figura 20: A quarta aproximação para a raiz de $f(x)$ já bem próxima de r para este exemplo

Se os números x_n ficam cada vez mais próximos de r à medida que n cresce, dizemos que a sequência converge para r e escrevemos

$$\lim_{n\to\infty} x_n = r.$$

Apesar da sequência de aproximações sucessivas convergir para a raiz desejada, como vimos no caso das funções do tipo ilustrado na Figura 20, em determinadas condições a sequência pode não convergir. Observe, por exemplo a situação mostrada na Figura 21. Podemos notar que x_2 é uma aproximação pior do que x_1.

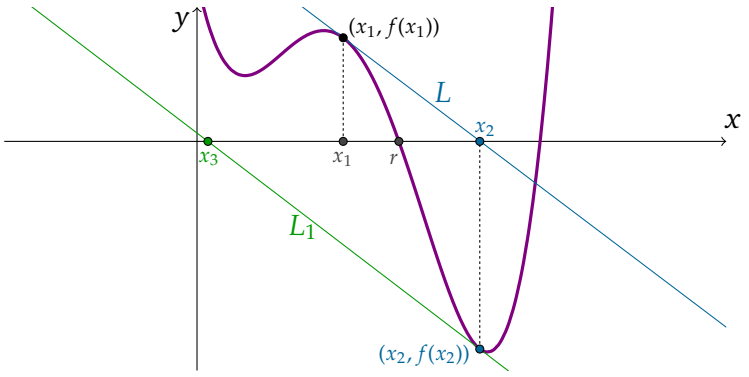

Figura 21: A aproximação x_2 é pior do que a aproximação x_1. E a aproximação x_3 é pior do que a aproximação x_2.

Nesta situação é necessário escolher uma nova aproximação inicial x_1.

Exemplo 7.1. Este exemplo foi usado pelo próprio Newton, segundo [13], para ilustrar o método. Começando com $x_1 = 2$, encontre a terceira aproximação x_3 para a raiz da equação $x^3 - 2x - 5 = 0$.

Solução: Vamos aplicar o método de Newton com

$$f(x) = x^3 - 2x - 5 \qquad \text{e} \qquad f'(x) = 3x^2 - 2.$$

Em [13] é colocado que Newton escolhe $x_1 = 2$, após alguns experimentos, pois $f(1) = -6$, $f(2) = -1$ e $f(3) = 16$. Assim a equação (7.1), fica

$$x_{n+1} = x_n - \frac{x_n^3 - 2x_n - 5}{3x_n^2 - 2}.$$

Para $n = 1$ temos

$$\begin{aligned} x_2 &= x_1 - \frac{x_1^3 - 2x_1 - 5}{3x_1^2 - 2} \\ &= 2 - \frac{2^3 - 2(2) - 5}{3(2)^2 - 2} = 2,1. \end{aligned}$$

E com $n = 2$ obtemos

$$\begin{aligned} x_3 &= x_2 - \frac{x_2^3 - 2x_2 - 5}{3x_2^2 - 2} \\ &= 2,1 - \frac{2,1^3 - 2(2,1) - 5}{3(2,1)^2 - 2} \approx 2,0946. \end{aligned}$$

Vamos dar um exemplo, de um resultado preciso até a oitava casa decimal. Este exemplo também foi extraído de [13].

Exemplo 7.2. Use o método de Newton para encontrar $\sqrt[6]{2}$ correta até a oitava casa decimal.

Solução: Note que encontrar a raiz de $\sqrt[6]{2}$ é equivalente a determinar a raiz positiva de $x^6 - 2 = 0$, assim podemos tomar $f(x) = x^6 - 2$ e $f'(x) = 6x^5$. Com isso, a equação (7.1) fica

$$x_{n+1} = x_n - \frac{x_n^6 - 2}{6x_n^5}.$$

Escolhendo $x_1 = 1$ como a aproximação inicial, obtemos:

$$x_2 \approx 1,16666667$$
$$x_3 \approx 1,12644368$$
$$x_4 \approx 1,12249707$$
$$x_5 \approx 1,12246205$$
$$x_6 \approx 1,12246205.$$

Uma vez que x_5 e x_6 são iguais até a oitava casa decimal, concluímos que

$$\sqrt[6]{2} \approx 1,12246205.$$

até a oitava casa decimal.

Agora, vamos aplicar o método de Newton para uma equação de quinto grau não resolúvel por radicais.

Exemplo 7.3. Vamos determinar uma raiz entre 0 e 1 para a equação $x^5 - 6x + 3 = 0$.

Solução: Primeiro observamos que existe uma raiz entre 0 e 1. Tomando

$$f(x) = x^5 - 6x + 3 \qquad \text{e} \qquad f'(x) = 5x^4 - 6,$$

percebemos que $f(0) = 3$ e $f(1) = -2$, logo existe uma raiz entre 0 e 1. Escrevendo a equação (7.1) temos

$$x_{n+1} = x_n - \frac{x_n^5 - 6x_n + 3}{5x_n^4 - 6}.$$

Escolhendo $x_1 = 0$ como a aproximação inicial, obtemos:

$$x_2 = 0,5$$
$$x_3 \approx 0,505494505495$$
$$x_4 \approx 0,505501230395$$
$$x_5 \approx 0,505501230405.$$

E assim conseguindo uma precisão de nove casas decimais na quinta aproximação.

7.2.1 *Características do Método de Newton*

O método de Newton que descrevemos anteriormente, pode ser apresentado por outra perspectiva. Vamos mostrar uma apresentação do método feita em [8], baseada

nos *polinômios de Taylor* (podemos encontrar mais sobre o polinômio de Taylor em [13]) e produz não só o método, mas também um limite para o erro da aproximação.

Em linhas gerais, podemos dizer que o *polinômio de Taylor* é uma aproximação para uma função $f(x)$ quando $x = a$. Isto é feito tentando encontrar aproximações melhores com polinômios de graus maiores. Assim, procuramos por um polinômio de grau n

$$T_n(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + c_3(x-a)^3 + \cdots + c_n(x-a)^n$$

tal que T_n e suas n primeiras derivadas tenham os mesmos valores em $x = a$ como f e suas n primeiras derivadas. E isto é satisfeito diferenciando[2] repetidamente e fazendo $x = a$ e $c_0 = f(a)$, $c_1 = f'(a)$, $c_2 = \frac{1}{2}f''(a)$, e em geral

$$c_k = \frac{f^{(k)}(a)}{k!},$$

onde $k! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot \cdots \cdot k$. O polinômio resultante

$$T_n(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n$$

é chamado de polinômio de Taylor de grau n de f centrado em a.

Suponhamos $f \in C^2[a,b]$[3]. Tomando $p_0 \in [a,b]$ uma aproximação para p tal que $f'(p_0) \neq 0$ e $|p - p_0|$ seja "pequeno". Considerando o primeiro termo para a expansão do polinômio de Taylor para $f(x)$ sobre p_0 e fazendo $x = p$, temos

$$f(p) = f(p_0) + (p-p_0)f'(p_0) + \frac{(p-p_0)^2}{2}f''(\xi(p)),$$

onde $\xi(p)$ se situa entre p e p_0. Desde $f(p) = 0$, esta equação dá

$$0 = f(p_0) + (p-p_0)f'(p_0) + \frac{(p-p_0)^2}{2}f''(\xi(p)).$$

O método de Newton é obtido assumindo-se que uma vez que $|p - p_0|$ é suficientemente pequeno, envolvendo o termo $(p-p_0)^2$ é muito menor, assim

$$0 \approx f(p_0) + (p-p_0)f'(p_0).$$

2 Estamos dizendo aqui que iremos derivar a função f várias vezes. Já dissemos antes que a derivada de f é f'. A derivada de f' é denotada de f'' e ao continuarmos este processo n vezes a enésima derivada é denotada por $f^{(n)}$.

3 $C^2[a,b]$ é conjunto de todas as funções com derivadas de segunda ordem, ou seja, todas as funções f' e f'' contínuas no intervalo $[a,b]$

Resolvendo para p obtemos

$$p \approx p_0 - \frac{f(p_0)}{f'(p_0)} \equiv p_1.$$

Isso prepara o terreno para o método de Newton, que começa com uma aproximação inicial p_0 e gera a sequência de $\{p_n\}_{n=1}^{\infty}$, por

$$p_n = p_{n-1} - \frac{f(p_{n-1})}{f'(p_{n-1}}, \qquad \text{for } n \geq 1. \tag{7.2}$$

Com isto podemos estabelecer o seguinte algoritmo (extraído de [8]) para encontrar uma solução para $f(x) = 0$ dado um p_0 aproximação inicial:

ENTRADA aproximação inicial p_0; tolerância TOL; número máximo de iterações N_0.
SAÍDA solução p ou mensagem de falha aproximada.
Passo 1 Fazer $i = 1$.
Passo 2 Enquanto $i \leq N_0$ fazer *Passos 3-6*.
Passo 3 Fazer $p = p_0 - f(P_0)/f'(p_0)$. (*Calcule* p_i.)
Passo 4 Se $|p - p_0| < TOL$, em seguida,
SAÍDA (p); (*O procedimento foi bem sucedido*.)
PARE.
Passo 5 Fazer $i = i + 1$.
Passo 6 Fazer $p_0 = p$. (*Atualiza* p_0.)
Passo 7 **SAÍDA** ('O método falhou após iterações N_0, N_0 =', N_0);
(*O procedimento foi bem sucedido*.)
PARE.

Desta maneira, selecionamos uma tolerância $\varepsilon > 0$, e construímos $p_1, \ldots, p_N$ até

$$|p_N - p_{N-1}| < \varepsilon, \tag{7.3}$$

$$\frac{|p_N - p_{N-1}|}{|p_N|} < \varepsilon, \qquad p_N \neq 0, \tag{7.4}$$

ou

$$|f(p_N)| < \varepsilon \tag{7.5}$$

A forma da desigualdade (7.3) é utilizada no *Passo 4* do algoritmo. Note-se que nenhuma das desigualdades (7.3), (7.4) ou (7.5) possuem informações precisas sobre o erro real $|p_N - p|$.

O método de Newton é uma técnica iteração funcional com $p_n = g(p_{n-1})$, para os quais

$$g(p_{n-1}) = p_{n-1} - \frac{f(p_{n-1})}{f'(p_{n-1})}, \qquad \text{for } n \geq 1.$$

É evidente a partir da equação (7.2) que o método de Newton não pode ser continuado se $f'(p_{n-1}) = 0$ para algum n. Na verdade, vamos ver que o método é mais eficaz quando f' é delimitada para uma distância próxima a zero de p.

7.2.2 *Convergência Usando o Método de Newton*

Partindo do polinômio de Taylor introduzido anteriormente, o método de Newton salienta a importância de uma aproximação inicial precisa. O pressuposto fundamental é que o termo envolvendo $(p - p_0)^2$ é, por comparação com $|p - p_0|$, tão pequeno que pode ser excluído. Isto será claramente falso, a menos que p_0 seja uma boa aproximação para p. Se p_0 não é suficientemente próximo da raiz real, há pouca razão para suspeitar que o método de Newton irá convergir para a raiz. No entanto, em alguns casos, mesmo pobres aproximação inicial produzirá convergência.

O seguinte teorema de convergência para o método de Newton ilustra a teórica importância da escolha de p_0 e está demonstrado em [8].

Teorema 7.4. *Seja $f \in C^2[a, b]$. Se $p \in (a, b)$ é tal que $f(p) = 0$ e $f'(p) \neq 0$, então existe um $\delta > 0$ de tal forma que o método de Newton gera uma sequência $\{p_n\}_{n=1}^{\infty}$ convergindo para p para qualquer aproximação inicial $p_0 \in [p - \delta, p + \delta]$.*

O Teorema 7.4 estabelece que, sob suposições razoáveis, o método de Newton converge escolhendo uma aproximação inicial suficientemente precisa. Isto também implica que a constante k que circunda o derivado de g, e, consequentemente, indica a velocidade de convergência do método, diminui para 0, quando o processo continua. Este resultado é importante para a teoria do método de Newton, mas raramente é aplicada na prática porque não nos diz como determinar δ.

Numa aplicação prática, uma aproximação inicial é selecionada e aproximações sucessivas são geradas pelo método de Newton. Estas geralmente convergirão rapidamente para a raiz, ou ficará claro que a convergência é improvável.

A

APÊNDICE A

A.1 COMO BHASKARA RESOLVIA AS EQUAÇÕES QUADRÁTICAS

Este exemplo foi retirado de [12].

Um método geral era enunciado para um problema escrito na forma padrão:

(I) "De uma quantidade retiramos ou adicionamos a sua raiz multiplicada por um coeficiente e a soma ou a diferença é igual a um número dado."

A quantidade citada é um quadrado e a raiz desse quadrado é a incógnita. Esse é um enunciado retórico que, traduzido em nossa notação, seria uma equação geral como $x^2 \pm bx = c$. O método de resolução consistia em reduzir o problema a uma igualdade, ou seja, sem o termo quadrado. Isso era feito por meio da técnica de "eliminação do termo médio":

(II) "Seja uma igualdade contendo a quantidade desconhecida, seu quadrado etc. Se temos os quadrados da quantidade desconhecida etc., em um dos membros multiplicamos os dois membros por um fator conveniente e somamos o que é necessário para que o membro das quantidades desconhecidas tenha uma raiz; igualando, em seguida, essa raiz à do membro das quantidades conhecidas, obtemos o valor da quantidade desconhecida."

Observamos que se concebia, de modo retórico, uma igualdade entre dois membros, sem utilização do sinal de igual: a igualdade entre um membro contendo a quantidade desconhecida (e o seu quadrado) e outro membro contendo as quantidades conhecidas. O primeiro membro deve ser escrito de modo a possuir uma raiz, ou seja, deve ser reescrito como um quadrado, o que se obtém pelas seguin-

tes especificações:

(III) "É por unidades iguais a quatro vezes o número de quadrados que é preciso multiplicar os dois membros; e é a quantidade igual ao quadrado do número primitivo de quantidades desconhecidas simples que é preciso adicionar." Temos, assim, a condição requerida em (II) de que o membro das quantidades desconhecidas tenha uma raiz. Trata-se do método que conhecemos hoje como "completar o quadrado".

Tradução do método de Bhaskara em nossa notação
Para resolver a equação $ax^2 + bx = c$:

- Multiplicamos ambos os lados por 4a, obtendo $4a^2x^2 + 4abx = 4ac$.
- Em seguida, adicionamos b^2 a ambos os lados, $4a^2x^2 + 4abx + b^2 = 4ac + b^2$.
- Agora podemos reescrever essa igualdade como $(2ax + b)^2 = 4ac + b^2$ e o membro contendo as quantidades desconhecidas possui uma raiz. Tomamos, então, a raiz quadrada para obter:

$$2ax + b = \pm\sqrt{4ac + b^2} \quad \text{e} \quad x = \frac{\pm\sqrt{4ac + b^2} - b}{2a}.$$

BIBLIOGRAFIA

[1] Jörg Bewersdorff e Traduzido em Inglês por David Kramer, *Galois Theory for Beginners: A Historical Perspective*, Student Mathematical Library, vol. 35, American Mathematical Society - AMS, Providence, Rhode Island, 2006, ISBN 978-0-8218-3817-4.

[2] Carl Benjamin Boyer e Traduzido em Português por Elza F. Gomide, *História da Matemática*, Edgard Blücher Ltda (obra publicada com a colaboração da Universidade de São Paulo), 1974.

[3] Jorge Delgado, Katia Frensel e Lhaylla Crissaff, *Geometria Analítica*, 1ª ed., Coleção PROFMAT, Sociedade Brasileira de Matemática - SBM, Rio de Janeiro, 2013, ISBN 978-85-8337-009-3.

[4] John B Fraleigh, *A First Course in Abstract Algebra*, 7ª ed., Pearson, 2003, ISBN 0-20176-390-7.

[5] Abramo Hefez e Cecília de Souza Fernandez, *Introdução à Álgebra Linear*, 1ª ed., Coleção PROFMAT, Sociedade Brasileira de Matemática - SBM, Rio de Janeiro, 2012, ISBN 978-85-85818-61-6.

[6] Israel Nathan Herstein, Traduzido em Português por Adalberto P. Bergamasco e L. H. Jacy Monteiro, *Tópicos de Álgebra*, Polígono S. A. (obra publicada com a colaboração da Universidade de São Paulo), São Paulo, 1970.

[7] Gelson Iezzi, Osvaldo Dolce e Antonio Machado, *Matemática e Realidade: 9° Ano*, 6ª ed., Atual, São Paulo, 2009, ISBN 978-85-357-1070-0.

[8] Richard L. Burden e J. Douglas Faires, *Numerical Analysis*, 9ª ed., Cengage Learning, Boston, Massachusetts, 2005, ISBN 0-538-73351-9.

[9] Elon Lages Lima, *Curso de Análise*, 8ª ed., vol. 1, Livros Técnicos e Científicos - LTC and Sociedade Brasileira de Matemática - SBM, Rio de Janeiro, 1976, ISBN 85-244-0047-1.

[10] Elon Lages Lima, *Números e Funções Reais*, 1ª ed., Coleção PROFMAT, Sociedade Brasileira de Matemática - SBM, Rio de Janeiro, 2013, ISBN 978-85-85818-81-4.

[11] Elon Lages Lima, Paulo Cezar Pinto Carvalho, Eduardo Wagner e Augusto César Morgado, *A Matemática do Ensino Médio*, 10ª ed., Coleção do Professor de Matemática, vol. 1, Sociedade Brasileira de Matemática - SBM, Rio de Janeiro, 2000, ISBN 978-85-85818-83-8.

[12] Tatiana Roque, *História da Matemática - Uma Visão Crítica, Desfazendo Mitos e Lendas*, 1ª ed., Zahar, Rio de Janeiro, 2012, ISBN 978-85-37808-88-7.

[13] James Stewart, *Cálculo*, 4ª ed., vol. 1, Pioneira Thomson Learning, São Paulo, 2003, ISBN 85-221-0265-X.

ÍNDICE

LISTA DE FIGURAS

LISTA DE TABELAS