Fatoração de números inteiros usando curvas elíticas

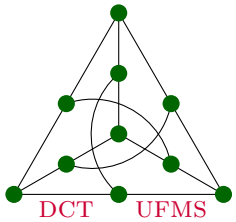
Celso Cardoso

Dissertação de Mestrado

Orientação: Profa. Dra. Elisabete Sousa Freitas

Área de Concentração: Ciência da Computação

Dissertação apresentada ao Departamento de Computação e Estatística da Universidade Federal de Mato Grosso do Sul como parte dos requisitos para a obtenção do título de mestre em Ciência da Computação.



Departamento de Computação e Estatística
Centro de Ciências Exatas e Tecnologia
Universidade Federal de Mato Grosso do Sul
Maio de 2003

Fatoração de números inteiros usando curvas elíticas

Este exemplar corresponde à redação final
da dissertação de mestrado devidamente
corrigida e defendida por Celso Cardoso
e aprovada pela comissão julgadora.

Campo Grande/MS, 12 de maio de 2003.

Banca Examinadora:

- Profa. Dra. Elisabete Sousa Freitas (orientadora) (DMT/CCET/UFMS)
- Prof. Dr. Edson Norberto Cáceres (DCT/CCET/UFMS)
- Prof. Dr. José Gilvan de Oliveira (DMAT/CCE/UFES)

Aos meus pais, Luiz e Deolinda;
À minha esposa, Sônia;
Aos meus filhos, Lívia, Matheus, Rafael e Ricardo.

Agradecimentos

À minha orientadora, Profa. Elisabete Sousa Freitas, pela paciência, pela compreensão e pela sua dedicação na orientação deste trabalho.

Aos colegas do Departamento de Matemática da Universidade Federal de Mato Grosso do Sul que acreditaram em mim, permitiram meu afastamento das atividades docentes para cursar o Mestrado e me incentivaram durante o curso. Em particular, agradeço aos colegas João Carlos, Paulo, Sônia e Wânia, pelos comentários e sugestões.

Aos professores do Departamento de Computação e Estatística da Universidade Federal de Mato Grosso do Sul que, de alguma forma, me ajudaram. Em especial, ao professor Marcelo Ferreira Siqueira, pelos comentários e sugestões.

Aos professores do curso que pela sua competência, dedicação e seriedade reacenderam meu gosto pelos estudos.

Ao Prof. Sergio Roberto de Freitas (in memoriam) pela sua ajuda nas minhas dúvidas com o Latex.

Aos colegas do mestrado pelo companheirismo e alegria, e cuja seriedade nos estudos sempre me serviram como estímulo; em especial, ao colega Amaury Antônio de Castro Junior.

Ao amigo João Carlos pelo constante incentivo.

Ao professor Horácio Braga pela ajuda nas correções da introdução do trabalho.

À minha família, pela compreensão da ausência; em particular, à minha esposa Sônia, por sua ajuda e sua paciência nos momentos mais difíceis.

Resumo

O problema da fatoração inteira tem obtido considerável atenção por sua utilização em sistemas criptográficos modernos que têm sua segurança baseada na dificuldade de fatorar números grandes.

Neste trabalho, apresentamos a descrição de um método de fatoração de números inteiros, o Método das Curvas Elíticas (Elliptic Curve Method - ECM) devido a H. W. Lenstra [Len87], que usa curvas elíticas. Ele é baseado num outro método de fatoração, o método $p-1$ de Pollard [Pol74]. O método de Pollard utiliza a estrutura do grupo multiplicativo $\mathbb{Z}_p^*$, enquanto o ECM utiliza a estrutura de grupo dos pontos de uma curva elítica.

Abstract

The Integer Factoring Problem has obtained considerable attention for its utilization in modern cryptographic systems which have its security based on the difficulty of factoring large numbers.

In this work, we present the description of a method for integer factorization, the Elliptic Curve Method - ECM, invented by H. W. Lenstra [Len87], which uses elliptic curves. It is based on another method for integer factorization, the Pollard $p-1$ method [Pol74]. The Pollard $p-1$ method uses the structure of the multiplicative group $\mathbb{Z}_p^*$, while the Elliptic Curve Method uses the group structure of the points of an elliptic curve.

Conteúdo

Capítulo 1

Introdução

Fatorar números inteiros é um problema aritmético tão simples de compreender que é abordado já no ensino fundamental. No entanto, fatorar números grandes (com mais de 100 dígitos decimais) é, em geral, um problema difícil. Existem números grandes que são fáceis de fatorar, por exemplo 100! e 10^{100}, mas, ao somarmos 1 a cada um deles, a fatoração torna-se difícil.

Um problema intimamente relacionado com o da fatoração é o de saber-se se um número inteiro é primo ou composto. Existem testes que nos permitem verificar se um número inteiro é composto ou provavelmente primo. Isto é relativamente fácil e rápido. Podemos, inclusive, saber, com certeza, que um número é composto, sem conseguir fatorá-lo. Há, também, testes que nos permitem saber se um número é realmente primo; são os chamados testes de primalidade. Não são, em geral, tão rápidos quanto os citados anteriormente (em agosto de 2002, foi publicado o artigo *Primes is in P*, por Manindra Agrawal, Neeraj Kayal e Nitin Saxena [AKS02], onde é apresentado um algoritmo determinístico que verifica, em tempo polinomial, se um número n é primo ou composto). Dado um número inteiro $N \geq 2$, o que fazemos, na prática, é, em primeiro lugar, verificar, rapidamente, se N é composto ou provavelmente primo. Se N for provavelmente primo, aplicamos um teste de primalidade para verificar se ele é realmente primo. Caso N seja composto, tentamos fatorá-lo, usando algum dos métodos de fatoração existentes. Um método de fatoração não dará, em geral, a fatoração completa de um número composto N, mas um fator não trivial de N, isto é, um fator d de N tal que $1 < d < N$. Assim, diremos, às vezes, por abuso de linguagem, que achar um fator primo de N é fatorar N. Por aplicações repetidas de um método de fatoração, podemos obter a fatoração completa de um número composto N.

O método mais natural para obter um fator primo de um número inteiro $N \geq 2$ é o método das divisões por tentativas. Tal método consiste em

dividir o número N por cada um dos números inteiros de 2 até $N-1$ (na verdade basta dividir pelos primos de 2 até $\lfloor\sqrt{N}\rfloor$). Se algum destes números dividir N, então encontramos um fator não trivial de N, e N será composto; caso contrário N será primo. Este método testa a primalidade de N e já apresenta um fator não trivial de N, no caso dele ser composto. Este método certamente achará um fator de N, porém ele não será eficiente para números grandes, pois o seu tempo de execução depende do tamanho de N. Para um número natural N da ordem de 10^{100}, por exemplo, precisaríamos executar em torno de 10^{50} operações de divisão. Supondo que cada operação demore 1 microssegundo, o algoritmo poderia demorar aproximadamente 3×10^{27} bilhões de anos.

O problema da fatoração inteira tem obtido considerável atenção desde que Ron Rivest, Adi Shamir e Len Adleman [RSA78] apresentaram o sistema criptográfico RSA, cuja segurança depende da dificuldade de fatorar grandes números. Este fato, junto com a crescente viabilidade dos computadores modernos, tem levado a um grande número de novos algoritmos de fatoração de inteiros.

Os métodos modernos de fatoração são muito mais rápidos que o método das divisões por tentativas. É claro que só tem sentido utilizar tais métodos quando se tem um computador como ferramenta. Para fatorar um número usando somente lápis e papel, a melhor maneira é usando o método das divisões por tentativas. Dentre os métodos mais poderosos usados hoje em dia estão o Crivo do Corpo de Números (Number Field Sieve - NFS) [LLMP89] e o Método das Curvas Elíticas (Elliptic Curve Method - ECM) [Len87].

Curvas elíticas têm sido estudadas em Teoria dos Números e Geometria Algébrica há aproximadamente 100 anos, e hoje existe uma grande quantidade de textos escritos sobre este assunto. Originalmente estudadas por razões puramente teóricas, as curvas elíticas têm sido utilizadas, recentemente, em planejamento de algoritmos para fatoração de inteiros, testes de primalidade e em criptografia de chave pública.

O objetivo deste trabalho é a descrição do Método das Curvas Elíticas. Antes de apresentá-lo, revisamos, no capítulo 2, os conceitos e resultados básicos da Teoria dos Números e Álgebra necessários para a compreensão do Método das Curvas Elíticas. No capítulo 3, estudamos a estrutura de grupo dos pontos de uma curva elítica, que será utilizada no método de fatoração ECM. No capítulo 4, descrevemos os métodos de fatoração $p-1$ de Pollard e o Método das Curvas Elíticas de Lenstra.

Capítulo 2

Conceitos e resultados básicos

Neste capítulo apresentaremos, de maneira sucinta, os conceitos e resultados básicos de Teoria dos Números e Álgebra necessários para a compreensão do Método de Fatoração de Números Inteiros de Lenstra, chamado de Método das Curvas Elíticas (veja [Gon79], [Hef93] e [MC00]).

2.1 Grupos e anéis

Definição 2.1.0.1 Seja G um conjunto munido de uma operação binária *,

$$\begin{aligned} * : G \times G &\to G \\ (x, y) &\mapsto x * y. \end{aligned}$$

Dizemos que o par $(G, *)$ é um grupo, quando são válidas as seguintes propriedades:

G_1 **(Associatividade)** Quaisquer que sejam a, b e c em G tem-se que $(a * b) * c = a * (b * c)$;

G_2 **(Existência de elemento neutro)** Existe um elemento e em G tal que $a * e = e * a = a$, $\forall a \in G$;

G_3 **(Existência de elemento simétrico)** Para cada $a \in G$, existe $b \in G$ tal que $a * b = b * a = e$.

Observação: Em um grupo $(G, *)$, o elemento neutro e o inverso de cada elemento são univocamente determinados. Quando usamos a notação multiplicativa $(* = \cdot)$, o elemento neutro e é denotado por 1 e chamado de elemento identidade (propriedade G_2); o elemento b, simétrico de a, é denotado por a^{-1} e chamado de inverso de a (propriedade G_3). Na notação

aditiva ($* = +$), o elemento neutro é denotado por 0 e o elemento b, que aparece em G_3, é denotado por $-a$ (oposto de a).
Se, além destas propriedades, vale G_4, a seguir, dizemos que $(G, *)$ é um grupo abeliano.

G_4 **(Comutatividade)** Quaisquer que sejam $a, b \in G$, tem-se $a * b = b * a$.

Por questões de simplicidade, muitas vezes, costumamos nos referir a um grupo $(G, *)$ simplesmente por G.

Definição 2.1.0.2 Dizemos que um grupo G é finito, quando ele tem um número finito de elementos; caso contrário, dizemos que G é um grupo infinito. Se um grupo G é finito, chamamos de ordem de G, o número de elementos de G, e denotamos tal número por $|G|$ ou $\#G$.

Exemplo 2.1.0.1 Seja $G_1 = \{e, a, b, c\}$ munido da operação * definida pela tabela abaixo.

$*$	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

G_1 é um grupo abeliano finito com 4 elementos, e elemento neutro e.

Exemplo 2.1.0.2 Seja $G_2 = \{e, r, s, t, u, v\}$ munido da operação * definida pela tabela abaixo.

$*$	e	r	s	t	u	v
e	e	r	s	t	u	v
r	r	s	e	u	v	t
s	s	e	r	v	t	u
t	t	v	u	e	r	s
u	u	t	v	s	e	r
v	v	u	t	r	s	e

G_2 é um grupo não-abeliano finito com 6 elementos, e elemento neutro e.

Exemplo 2.1.0.3 $(\mathbb{Z}, +)$ é um grupo abeliano de ordem infinita, onde $\mathbb{Z}$ é o conjunto dos números inteiros $\{..., -3, -2, -1, 0, 1, 2, 3, ...\}$ e $+$ é a operação de adição usual em $\mathbb{Z}$.

Definição 2.1.0.3 Sejam G um grupo e $H \subseteq G$ tal que $H \neq \emptyset$. Dizemos que H é um subgrupo de G, se H for, ele próprio, um grupo com a operação de G restrita à H.

Definição 2.1.0.4 Seja $(G, \cdot)$ um grupo com elemento neutro 1. Dado $x \in G$, definimos x^n $(n \in \mathbb{Z})$ da seguinte maneira:

$$x^n = \begin{cases} 1, & \text{se} \quad n = 0 \\ x^{n-1} \cdot x, & \text{se} \quad n > 0 \\ (x^{-n})^{-1}, & \text{se} \quad n < 0 \end{cases}$$

Observação: Na definição acima usamos a notação multiplicativa. Quando $n > 0$, $x^n = \underbrace{x \cdot \ldots \cdot x}_{n \text{ vezes}}$. Na notação aditiva escrevemos $n \cdot x = \underbrace{x + \ldots + x}_{n \text{ vezes}}$.

Proposição 2.1.0.1 Seja $(G, \cdot)$ um grupo finito com elemento neutro 1. Dado $a \in G$, considere $H = \{a^n \mid n \in \mathbb{Z}\} \subseteq G$. Então,

(i) H é um subgrupo de G.

(ii) A ordem de H é igual ao menor inteiro positivo k tal que $a^k = 1$.

(iii) $H = \{1, a, a^2, \ldots, a^{k-1}\}$, onde k é a ordem de H.

Dizemos que o elemento a é um *gerador* de H e que k é a *ordem* de a. Denotamos a ordem de a por $o(a)$. Um subgrupo qualquer de G que admite um elemento gerador é chamado subgrupo cíclico. Em particular, se G admite um elemento gerador, ele é chamado de grupo cíclico.

Proposição 2.1.0.2 Sejam $(G, \cdot)$ um grupo finito e $a \in G$. Um número inteiro $t > 0$ satisfaz a equação $a^t = 1$ se, e somente se, $o(a)$ divide t.

Proposição 2.1.0.3 (Teorema de Lagrange) Se G é um grupo finito e H é um subgrupo de G, então $|H|$ é um divisor de $|G|$, isto é, a ordem de H é um divisor da ordem de G.

Definição 2.1.0.5 Seja A um conjunto onde estão definidas duas operações binárias

$$\begin{array}{c} + : A \times A \to A \\ (x, y) \mapsto x + y \end{array} \quad \text{e} \quad \begin{array}{c} \cdot : A \times A \to A \\ (x, y) \mapsto x \cdot y \end{array}$$

chamadas de adição e de multiplicação, respectivamente.

Dizemos que a terna $(A, +, \cdot)$ é um anel comutativo com unidade, ou simplesmente anel, quando são válidas as seguintes propriedades:

A_1 **(Associatividade da Adição)** Quaisquer que sejam $a, b, c \in A$, tem-se que $(a + b) + c = a + (b + c)$;

A_2 **(Comutatividade da Adição)** Quaisquer que sejam $a, b \in A$, tem-se que $a + b = b + a$;

A_3 **(Existência de elemento neutro para a Adição)** Existe um elemento em A, denotado por 0, tal que $a + 0 = 0 + a = a$, para todo $a \in A$;

A_4 **(Existência de elemento simétrico para cada $a \in A$)** Para cada $a \in A$, existe $b \in A$ tal que $a + b = b + a = 0$;

M_1 **(Associatividade da Multiplicação)** Quaisquer que sejam $a, b, c \in A$, tem-se que $(a \cdot b) \cdot c = a \cdot (b \cdot c)$;

M_2 **(Comutatividade da Multiplicação)** Quaisquer que sejam $a, b \in A$, tem-se que $a \cdot b = b \cdot a$;

M_3 **(Existência de elemento neutro para a Multiplicação)** Existe um elemento em A, denotado por 1, com $1 \neq 0$, tal que $a \cdot 1 = a$ qualquer que seja $a \in A$;

AM **(Distributividade da Multiplicação em relação à Adição)** Quaisquer que sejam $a, b, c \in A$, tem-se que $(a + b) \cdot c = a \cdot c + b \cdot c$.

Definição 2.1.0.6 Dizemos que um elemento a de um anel A é invertível, quando existe $b \in A$ tal que $a \cdot b = 1$. Tal elemento b é único, é denominado inverso de a e denotado por a^{-1}.

Definição 2.1.0.7 Dizemos que um anel A é um corpo, quando todo elemento não nulo de A é invertível.

Assim, num corpo $(A, +, \cdot)$, além das propriedades de anel, vale a seguinte propriedade:

M_4 **(Existência de elemento inverso)** Qualquer que seja $a \in A$, $a \neq 0$, existe $b \in A$ tal que $a \cdot b = 1$.

2.2 Os inteiros

Definição 2.2.0.8 Sejam a e b inteiros. Dizemos que a *divide* b, se existe um inteiro c tal que $b = a \cdot c$. Quando a *divide* b, dizemos também que a é

divisor de b, ou a *é fator de* b, ou b é múltiplo de a. Se a *divide* b, então denotamos isto por $a \mid b$.

Proposição 2.2.0.4 (Propriedades da divisibilidade) Para todos a, b, c e d em $\mathbb{Z}$, valem as seguintes propriedades:

(i) $a \mid 0$ e $a \mid a$;

(ii) Se $a \mid b$ e $b \mid c$, então $a \mid c$;

(iii) Se $a \mid b$ e $c \mid d$, então $a.c \mid b.d$;

(iv) Se $a \mid (b + c)$ e $a \mid b$, então $a \mid c$;

(v) Se $a \mid b$ e $a \mid c$, então $a \mid (b \cdot x + c \cdot y)$ para todos $x, y \in \mathbb{Z}$;

(vi) Se $a \mid b$ e $b \mid a$, então $a = \pm b$.

Proposição 2.2.0.5 (Algoritmo da divisão para inteiros) Se a e b são inteiros com $b \geq 1$, então existem inteiros q e r, únicos, tais que

$$a = q \cdot b + r, \text{onde } 0 \leq r < b.$$

Dizemos que q é o quociente e r o resto da divisão de a por b.

Chamamos de piso de um real n, o maior inteiro menor do que ou igual a n. Denotamos o piso de n por $\lfloor n \rfloor$. Dados dois inteiros a e b, denotamos por $a \bmod b$, o resto da divisão de a por b. Com estas notações, temos que $q = \left\lfloor \frac{a}{b} \right\rfloor$ e $r = a \bmod b$. Podemos, portanto, escrever

$$a = \left\lfloor \frac{a}{b} \right\rfloor b + a \bmod b.$$

Definição 2.2.0.9 Dizemos que um inteiro c é um *divisor comum* de dois inteiros a e b, se $c \mid a$ e $c \mid b$.

Definição 2.2.0.10 O *máximo divisor comum* de dois inteiros a e b (a e b não ambos nulos) é um inteiro positivo d, denotado por $d = mdc(a, b)$, que satisfaz as seguintes condições:

(i) d é um divisor comum de a e b;

(ii) Se $c \mid a$ e $c \mid b$, então $c \mid d$.

Definição 2.2.0.11 Dizemos que dois inteiros a e b são *primos entre si*, ou *primos relativos*, se $mdc(a, b) = 1$.

Proposição 2.2.0.6 Se a e b são números inteiros e $a = b \cdot q + r$, onde q e r são números inteiros, então $mdc(a, b) = mdc(b, r)$. Em particular, $mdc(a, b) = mdc(b, a \mod b)$.

Proposição 2.2.0.7 Sejam a e b números inteiros positivos.

(i) Se $mdc(a, b) = d$, então existem inteiros x e y tais que $a \cdot x + b \cdot y = d$.

(ii) $mdc(a, b) = 1$ se, e somente se, existem inteiros x e y tais que $a \cdot x + b \cdot y = 1$.

Definição 2.2.0.12 Dizemos que um número inteiro c é um *múltiplo comum* de dois números inteiros a e b, se $a \mid c$ e $b \mid c$.

Definição 2.2.0.13 O *mínimo múltiplo comum* de dois números inteiros a e b é um inteiro não negativo c, denotado por $mmc(a, b)$, que satisfaz as seguintes condições:

(i) c é um múltiplo comum de a e b;

(ii) Se $a \mid d$ e $b \mid d$, então $c \mid d$.

Definição 2.2.0.14 Dizemos que um inteiro $p \geq 2$ é um *número primo*, se os seus únicos divisores positivos são 1 e o próprio p; caso contrário dizemos que p é um *número composto*.

As proposições seguintes são propriedades bem conhecidas dos números primos.

Proposição 2.2.0.8 Sejam a, b e c inteiros tais que a e b são primos entre si.

(i) Se $b \mid a \cdot c$, então $b \mid c$.

(ii) Se $a \mid c$ e $b \mid c$, então $a \cdot b \mid c$.

Proposição 2.2.0.9 Se p é um número primo tal que $p \mid a \cdot b$, então $p \mid a$ ou $p \mid b$.

Proposição 2.2.0.10 (Teorema de Euclides) Existem infinitos números primos.

Proposição 2.2.0.11 (Teorema Fundamental da Aritmética) Todo inteiro $n \geq 2$ tem uma fatoração como produto de potências de primos,

$$n = p_1^{e_1} \cdot p_2^{e_2} \cdot \ldots \cdot p_k^{e_k},$$

onde $p_1, \ldots, p_k$ são primos distintos e $e_1, \ldots, e_k$ são inteiros positivos. Além disso, tal fatoração é única a menos da ordem dos fatores.

Definição 2.2.0.15 Para $n \geq 1$, seja $\phi(n)$ o número de inteiros do intervalo $[1, n]$ que são primos relativos com n. A função ϕ que a cada inteiro positivo n associa $\phi(n)$ é chamada de *função ϕ de Euler*.

Proposição 2.2.0.12 (Propriedades da função ϕ de Euler)

(i) Se p é primo, então $\phi(p) = p - 1$.

(ii) A função ϕ de Euler é multiplicativa, isto é, se $mdc(m, n) = 1$, então $\phi(mn) = \phi(m) \cdot \phi(n)$.

(iii) Se $n = p_1^{e_1} \cdot p_2^{e_2} \cdots p_k^{e_k}$ é a fatoração de n em potências de primos, então

$$\phi(n) = n\left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_2}\right)\cdots\left(1 - \frac{1}{p_k}\right).$$

2.3 Os inteiros módulo n

Definição 2.3.0.16 Seja n um inteiro positivo. Dizemos que dois inteiros a e b são *congruentes módulo n*, se a e b deixam o mesmo resto quando divididos por n. Se a e b são congruentes módulo n, escrevemos $a \equiv b \pmod{n}$.

Proposição 2.3.0.13 Tem-se que $a \equiv b \pmod{n}$ se, e somente se, $n \mid a - b$.

Proposição 2.3.0.14 (Propriedades da congruência) Para todos a, b, c, d, m e n em $\mathbb{Z}$ com $m \geq 1$ e $n > 1$, valem as seguintes propriedades:

(i) $a \equiv a \pmod{n}$;

(ii) Se $a \equiv b \pmod{n}$, então $b \equiv a \pmod{n}$;

(iii) Se $a \equiv b \pmod{n}$ e $b \equiv c \pmod{n}$, então $a \equiv c \pmod{n}$;

(iv) Se $a \equiv b \pmod{n}$ e $c \equiv d \pmod{n}$, então $a + c \equiv b + d \pmod{n}$;

(v) Se $a \equiv b \pmod{n}$ e $c \equiv d \pmod{n}$, então $ac \equiv bd \pmod{n}$;

(vi) Se $a \equiv b \pmod{n}$, então $a^m \equiv b^m \pmod{n}$.

Observa-se, pelas propriedades (i), (ii) e (iii), que, fixado um inteiro positivo n, a relação de congruência módulo n é uma relação de equivalência.

Definição 2.3.0.17 Dado $a \in \mathbb{Z}$, a classe de equivalência de a, denotada por $\overline{a} = \{x \in \mathbb{Z} \mid x \equiv a \pmod{n}\}$, chama-se *classe residual módulo n* do elemento a.

Proposição 2.3.0.15 Existem exatamente n classes residuais módulo n distintas, a saber, $\overline{0}, \overline{1}, \ldots, \overline{n-1}$.

Definição 2.3.0.18 O conjunto de todas as classes residuais módulo n chama-se conjunto dos *inteiros módulo n* e é denotado por $\mathbb{Z}_n$.

Freqüentemente, usamos a definição $\mathbb{Z}_n = \{0, 1, 2, \ldots, n-1\}$ que deve ser lida como equivalente à definição acima, $\mathbb{Z}_n = \{\overline{0}, \overline{1}, \overline{2}, \ldots, \overline{n-1}\}$, com 0 no lugar de $\overline{0}$, 1 no lugar de $\overline{1}$, etc. Cada classe é representada pelo seu menor elemento não negativo.
Em $\mathbb{Z}_n$, definimos as seguintes operações:
Adição: $\overline{a_1} + \overline{a_2} = \overline{a_1 + a_2}$
Multiplicação: $\overline{a_1} \cdot \overline{a_2} = \overline{a_1 \cdot a_2}$
Estas operações estão bem definidas, e $\mathbb{Z}_n$ munido destas operações é um anel.

Proposição 2.3.0.16 Seja $\overline{a} \in \mathbb{Z}_n$. Então $\overline{a}$ é invertível se, e somente se, $mdc(a, n) = 1$.

Seja $\mathbb{Z}_n^*$ o conjunto dos elementos invertíveis de $\mathbb{Z}_n$. Pela Proposição 2.3.0.16, $\mathbb{Z}_n^* = \{\overline{a} \in \mathbb{Z}_n \mid mdc(a, n) = 1\}$. Assim, a ordem de $\mathbb{Z}_n^*$, isto é, o número de elementos de $\mathbb{Z}_n^*$, é $\phi(n)$. É claro que $(\mathbb{Z}_n^*, \cdot)$ é um grupo. Em particular, se p é um inteiro primo, então $\mathbb{Z}_p^* = \{\overline{1}, \ldots, \overline{p-1}\}$, e $(\mathbb{Z}_p, +, \cdot)$ têm uma estrutura de corpo.

A seguir, enunciamos e demonstramos o Pequeno Teorema de Fermat, que será utilizado nos métodos de fatoração $p-1$ de Pollard e de Lenstra.

Proposição 2.3.0.17 (Pequeno Teorema de Fermat) Sejam p um número primo e a um número inteiro.

(i) Se $mdc(a, p) = 1$, então $a^{p-1} \equiv 1 \pmod{p}$.

(ii) Para qualquer inteiro a, $a^p \equiv a \pmod{p}$.

Prova:

(i) Consideremos o conjunto de números inteiros

$$\{a, 2a, \ldots, (p-1)a\}. \tag{2.1}$$

Nenhum dos elementos desse conjunto é congruente a zero módulo p, pois se assim o fosse, teríamos $p \mid xa$, com $1 \leq x \leq p-1$, e conseqüentemente $p \mid x$ ou $p \mid a$, o que não ocorre. Além disso, dois elementos quaisquer de 2.1 não são congruentes módulo p, pois se tivéssemos $xa \equiv ya \pmod{p}$, com $1 \leq x, y \leq p-1$, então, como $mdc(a,p) = 1$, teríamos $x \equiv y \pmod{p}$, o que não acontece, pois os elementos do conjunto

$$\{1, 2, \ldots, (p-1)\} \tag{2.2}$$

não são congruentes entre si, módulo p. Assim sendo, cada elemento de 2.1 é congruente a um único elemento de 2.2 . Temos, então, as $p-1$ congruências a seguir:

$$\begin{array}{ccc} a & \equiv & x_1 \pmod{p} \\ 2a & \equiv & x_2 \pmod{p} \\ \vdots & \vdots & \vdots \\ (p-1)a & \equiv & x_{p-1} \pmod{p} \end{array}$$

onde os x_i's são os elementos de 2.2, considerados numa certa ordem. Multiplicando, ordenadamente, todas as congruências acima, obtemos

$$a \cdot (2a) \cdot \ldots \cdot (p-1)a \equiv 1 \cdot 2 \cdot \ldots \cdot (p-1) \pmod{p},$$

isto é,

$$(p-1)!a^{p-1} \equiv (p-1)! \pmod{p}.$$

Como $mdc((p-1)!, p) = 1$, temos que $p \mid a^{p-1} - 1$, isto é, $a^{p-1} \equiv 1 \pmod{p}$.

(ii) Vimos no caso (i) que se $p \nmid a$, então $a^{p-1} \equiv 1 \pmod{p}$. Multiplicando ambos os termos dessa congruência por a, obtemos $a^p \equiv a \pmod{p}$. Se $p \mid a$, então $p \mid a^p$, donde segue que $p \mid (a^p - a)$, i.e., $a^p \equiv a \pmod{p}$.

Proposição 2.3.0.18 Seja p um número primo. Para todos r e s em $\mathbb{Z}$ tais que $r \equiv s \pmod{(p-1)}$ tem-se $a^r \equiv a^s \pmod{p}$, qualquer que seja o inteiro a. De outra forma, quando trabalha-se módulo um primo p, expoentes podem ser reduzidos módulo $p-1$.

Uma generalização do Teorema de Fermat é apresentada na próxima proposição.

Proposição 2.3.0.19 Seja $n \geq 2$ um inteiro.

(i) **(Teorema de Euler)** Se $a \in \mathbb{Z}$ e $mdc(a,n) = 1$ então, $a^{\phi(n)} \equiv 1 \pmod{n}$.

(ii) Se $n \in \mathbb{Z}$ e se $r \equiv s \pmod{\phi(n)}$, então $a^r \equiv a^s \pmod{n}$, para todo inteiro a. Em outras palavras, quando trabalhamos módulo n, os expoentes podem ser reduzidos módulo $\phi(n)$.

A definição de ordem de um elemento pode ser escrita na seguinte forma, quando consideramos o grupo $(\mathbb{Z}_n^*, \cdot)$.

Definição 2.3.0.19 Seja $\overline{a} \in \mathbb{Z}_n^*$. A *ordem* de $\overline{a}$, denotada por $o(\overline{a})$, é o menor inteiro positivo t tal que $a^t \equiv 1 \pmod{n}$.

Proposição 2.3.0.20 Se a ordem de $\overline{a} \in \mathbb{Z}_n^*$ é t e se $a^s \equiv 1 \pmod{n}$, então t divide s. Em particular, $t \mid \phi(n)$.

Exemplo 2.3.0.4 $\mathbb{Z}_{21}^* = \{\overline{1}, \overline{2}, \overline{4}, \overline{5}, \overline{8}, \overline{10}, \overline{11}, \overline{13}, \overline{16}, \overline{17}, \overline{19}, \overline{20}\}$. Observa-se que $\phi(21) = \phi(7) \cdot \phi(3) = 12 = |\mathbb{Z}_{21}^*|$. A ordem de cada elemento de $\mathbb{Z}_{21}^*$ está listada na tabela abaixo:

$\overline{a} \in \mathbb{Z}_{21}^*$	$\overline{1}$	$\overline{2}$	$\overline{4}$	$\overline{5}$	$\overline{8}$	$\overline{10}$	$\overline{11}$	$\overline{13}$	$\overline{16}$	$\overline{17}$	$\overline{19}$	$\overline{20}$
ordem de $\overline{a}$	1	6	3	6	2	6	6	2	3	6	6	2

Se $(\mathbb{Z}_n^*, \cdot)$ tem um elemento $\overline{a}$ de ordem $\phi(n)$, então $(\mathbb{Z}_n^*, \cdot)$ é um grupo cíclico gerado por $\overline{a}$. Neste caso, dizemos que $\overline{a}$ é um *elemento primitivo* de $(\mathbb{Z}_n^*, \cdot)$.

2.4 Característica de um Corpo

Definição 2.4.0.20 Sejam A e B anéis. Dizemos que uma função $f : A \longrightarrow B$ é um homomorfismo de anéis, ou simplesmente homomorfismo, se f satisfaz as seguintes propriedades:

(i) $f(a + a') = f(a) + f(a')$, para todos $a, a' \in A$;

(ii) $f(a \cdot a') = f(a) \cdot f(a')$, para todos $a, a' \in A$;

(iii) $f(1) = 1$.

Se f satisfaz as condições (i), (ii) e (iii) acima, então

(iv) $f(0) = 0$ e

(v) $f(-a) = -f(a)$ quaisquer que sejam $a, a' \in A$.

Além disso, vale a propriedade:

(vi) Se a tem inverso a^{-1} em A, então $f(a)$ tem inverso em B e $[f(a)]^{-1} = f(a^{-1})$.

Definição 2.4.0.21 Seja $f : A \longrightarrow B$ um homomorfismo de anéis. O núcleo de f, denotado por $ker(f)$, é o conjunto dos elementos a de A tais que $f(a) = 0$, isto é, $ker(f) = \{a \in A \mid f(a) = 0\}$.

Obviamente, este conjunto é sempre não vazio, uma vez que $f(0) = 0$.

A proposição a seguir apresenta uma condição necessária e suficiente para que um homomorfismo seja injetor.

Proposição 2.4.0.21 Um homomorfismo f é injetor se, e somente se, $ker(f) = \{0\}$.

Proposição 2.4.0.22 Sejam $\mathbb{Z}$ o anel dos inteiros com a adição e multiplicação usuais e A um anel qualquer. Então, a função $f : \mathbb{Z} \longrightarrow A$ dada por $f(n) = n \cdot 1_A$, onde 1_A é a unidade de A, é um homomorfismo. Além disso, f é o único homomorfismo de $\mathbb{Z}$ em A.

Seja f o homomorfismo dado na Proposição 2.4.0.22. Se f for injetor, temos que $ker(f) = \{0\}$; caso contrário, existe um $c \in ker(f)$ tal que $c \neq 0$. Mas $f(c) = 0$ implica em $f(-c) = -f(c) = -0 = 0$ e, portanto, existe um inteiro não negativo no núcleo de f.

Proposição 2.4.0.23 Se $f : \mathbb{Z} \longrightarrow A$ é um homomorfismo e m é o menor inteiro não negativo em $ker(f)$, então $ker(f) = m\mathbb{Z} = \{mk \mid k \in \mathbb{Z}\}$.

Proposição 2.4.0.24 Sejam K um corpo e $f : \mathbb{Z} \longrightarrow K$ uma função dada por $f(n) = n \cdot 1_K$. Se $ker(f) = m\mathbb{Z}$ e $m \neq 0$, então m é primo.

Vamos agora definir a característica de um corpo.

Definição 2.4.0.22 Seja K um corpo e $f : \mathbb{Z} \longrightarrow K$ dada por $f(n) = n \cdot 1_K$.

(i) Se $ker(f) = \{0\}$, dizemos que K tem característica 0;

(ii) Se $ker(f) = p\mathbb{Z}$, onde p é um número primo, dizemos que K tem característica p.

Capítulo 3

Curvas Elíticas

Neste capítulo apresentaremos, inicialmente, alguns conceitos básicos do estudo de curvas algébricas planas. Em seguida trataremos de interseções de curvas planas (teorema de Bezout). Veremos o conceito de plano projetivo e finalmente a definição de curvas elíticas.

3.1 Curvas Planas e Plano Projetivo

Definição 3.1.0.23 Sejam K um corpo e f um polinômio não constante em $K[x, y]$, o anel dos polinômios nas variáveis x e y. O conjunto dos pontos (x, y) tais que $x, y \in K$ será chamado de plano afim e denotado por $\mathbb{A}^2(K)$ ou, simplesmente, $\mathbb{A}^2$. Uma *curva algébrica plana* (ou *curva plana afim*) é o conjunto dos pares $(x, y) \in \mathbb{A}^2$ tais que $f(x, y) = 0$. O grau de uma curva C, denotado por grC, será, por definição, o grau do polinômio que a define.

Vamos apresentar alguns exemplos de equações que representam curvas algébricas planas. Consideremos $K = \mathbb{R}$, onde $\mathbb{R}$ é o corpo dos números reais. As equações do tipo $ax + by + c = 0$ representam retas em $\mathbb{A}^2$, onde a, b e c são constantes com a e b não ambas nulas. A equação $x^2 + y^2 - 1 = 0$ representa uma circunferência em $\mathbb{A}^2$, $2x^2 - 3y + 1 = 0$ uma parábola e $x^2 - y^2 = 1$ uma hipérbole. A curva dada pela equação $x^3 + y^2 = 3axy$, onde a é uma constante, chama-se Folium de Descartes e a curva dada pela equação $(x^2 + y^2)^3 = 16(x^2 - y^2)^2$ é uma rosácea de quatro pétalas.

Nos exemplos acima, consideramos $K = \mathbb{R}$. Se fizéssemos $K = \mathbb{Q}$, onde $\mathbb{Q}$ é o corpo dos números racionais, $K = \mathbb{Z}_p$ ou K igual a um outro corpo qualquer, teríamos também curvas algébricas planas. Não teríamos, em todos os casos, porém, a interpretação geométrica. Dependendo do corpo K que estamos considerando, uma equação pode representar o conjunto vazio.

Por exemplo, se $K = \mathbb{R}$, a equação $x^2+y^2 = 3$, representa uma circunferência.

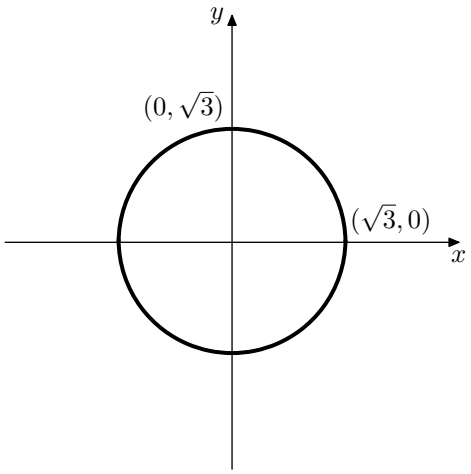


Figura 3.1

No entanto, se $K = \mathbb{Q}$, o conjunto dos pontos (x, y) que satisfazem essa mesma equação é o conjunto vazio. Em outras palavras, não existem pontos (x, y) com ambas as coordenadas racionais tais que $x^2+y^2 = 3$. Para mostrar isso, basta mostrar que não existem números inteiros x, y e w, primos entre si, tais que $x^2 + y^2 = 3w^2$. Suponhamos que existam inteiros x, y e w, sem fator comum, tais que $x^2 + y^2 = 3w^2$. Vamos mostrar que $3 \nmid x$ e $3 \nmid y$. Suponhamos, por absurdo, que $3 \mid x$. Então, como $x^2 + y^2 = 3w^2$, temos que $3 \mid y^2$ e, portanto, $3 \mid y$. Segue disso que $9 \mid (x^2+y^2)$, e daí temos que $9 \mid 3w^2$. Concluímos, então, que $3 \mid w$, o que é absurdo, já que x, y e w não têm fator comum. Portanto $3 \nmid x$. Por simetria, vemos que $3 \nmid y$. Assim, como x e y não são congruentes com 0 módulo 3, devemos ter $x, y \equiv 1$ ou 2, módulo 3, ou equivalentemente, $x, y \equiv \pm 1 \pmod 3$, o que acarreta em $x^2 + y^2 \equiv 2 \pmod 3$. Logo, $x^2 + y^2$ não é múltiplo de 3. Portanto, não existem racionais x e y tais que $x^2 + y^2 = 3$.

Agora, se $K = \mathbb{Z}_5$, é fácil ver, por simples substituição, que a curva dada pela equação $x^2 + y^2 = 3$ é formada pelos pontos $(2, 2), (2, 3), (3, 2), (3, 3)$ em $\mathbb{A}^2(\mathbb{Z}_5) = \mathbb{Z}_5 \times \mathbb{Z}_5$.

Vamos, agora, estudar o plano projetivo. Para motivar tal estudo, consideraremos o problema de encontrar os pontos de interseção de duas curvas algébricas planas. Sejam então, duas curvas planas $C_1 : f(x, y) = 0$ e $C_2 : g(x, y) = 0$ em $\mathbb{A}^2(K)$, onde f e g são polinômios de graus m e n, respectivamente. Quantos são os pontos de interseção de C_1 e C_2? Um teorema

que veremos mais adiante, o Teorema de Bezout, assegura que, desde que interpretemos adequadamente a pergunta, a resposta será $m.n$ pontos.

Consideremos $K = \mathbb{R}$. Começaremos estudando o caso em que C_1 e C_2 são retas, isto é, os polinômios que definem C_1 e C_2 são polinômios de grau 1.

Sabe-se da Geometria Euclidiana que dois pontos determinam uma única reta, a saber a reta que passa por eles. Sabe-se, também, que duas retas no plano euclidiano determinam um único ponto, a saber o ponto onde elas se interceptam; a menos que elas sejam paralelas. Então, no caso de duas retas paralelas, que são curvas dadas por polinômios de grau 1, não teremos um ponto de interseção. Vamos, então, considerar um novo conjunto, chamado de plano projetivo, que contenha o plano euclidiano, preserve a propriedade "dois pontos determinam uma única reta" e no qual "duas retas quaisquer (inclusive as retas paralelas) tenham exatamente um ponto de interseção". Quantos pontos extras precisaremos para que qualquer par de retas paralelas tenha um ponto de interseção? Seria suficiente usar um ponto extra P apenas e assumir que quaisquer duas retas paralelas se interceptem nele? A resposta é não; vejamos porque. Sejam L_1 e L_2 duas retas paralelas e P o ponto extra onde elas se interceptam. Do mesmo modo, sejam $L_1^{'}$ e $L_2^{'}$ duas retas paralelas, que se interceptam no ponto extra $P^{'}$. (Veja figura 3.2 abaixo.)

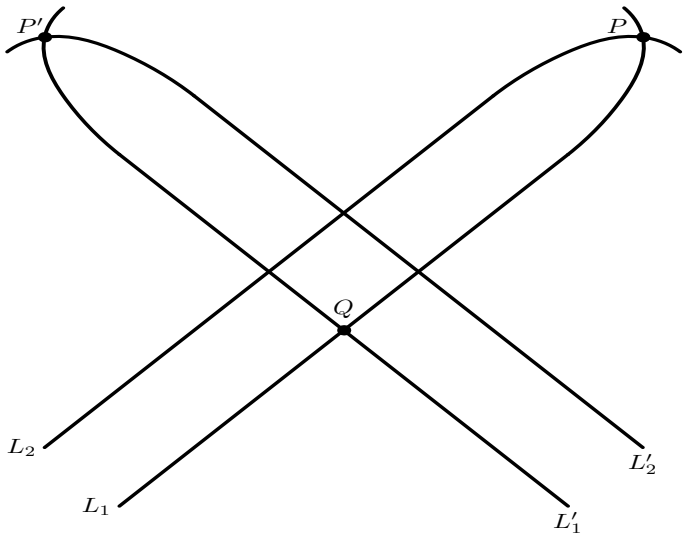


Figura 3.2

Suponhamos que L_1 e $L_1^{'}$ não são paralelas. Então L_1 e $L_1^{'}$ se interceptam em algum ponto ordinário Q. Mas como duas retas quaisquer só podem ter um

ponto em comum, os pontos P e $P^{'}$ devem ser distintos. Portanto, precisamos adicionar um ponto extra para cada direção distinta no plano ordinário e determinar que uma reta L consiste de seus pontos usuais juntamente com o ponto extra determinado por sua direção. O plano euclidiano usual, $\mathbb{A}^2(\mathbb{R})$, acrescido destes pontos extras será chamado de plano projetivo e denotado por $\mathbb{P}^2(\mathbb{R})$. Assim $\mathbb{P}^2(\mathbb{R})$ será a união de $\mathbb{A}^2(\mathbb{R})$ com o conjunto das direções em $\mathbb{A}^2(\mathbb{R})$, isto é,

$$\mathbb{P}^2(\mathbb{R}) = \mathbb{A}^2(\mathbb{R}) \cup \{\text{conjunto das direções em } \mathbb{A}^2(\mathbb{R})\},$$

onde direção é uma noção não orientada.

Duas retas têm a mesma direção se, e somente se, elas são paralelas. Portanto, podemos definir uma direção como sendo uma classe de equivalência de retas paralelas, isto é, uma direção é uma coleção de todas as retas paralelas à uma reta dada. Os pontos extras em $\mathbb{P}^2(\mathbb{R})$, associados à direções, isto é, os pontos em $\mathbb{P}^2(\mathbb{R})$ que não estão em $\mathbb{A}^2(\mathbb{R})$, são freqüentemente chamados de pontos no infinito. Como indicado acima, uma reta em $\mathbb{P}^2(\mathbb{R})$ consiste de uma reta em $\mathbb{A}^2(\mathbb{R})$ junto com o ponto no infinito especificado por sua direção. A interseção de duas retas paralelas é o ponto no infinito que corresponde à sua direção comum. Finalmente, o conjunto de todos os pontos no infinito é, ele próprio, considerado como uma reta, denotado por L_∞ ou $\mathbb{P}^1(\mathbb{R})$, e a intersecção de qualquer outra reta L com L_∞ é o ponto no infinito que corresponde à direção de L. Com essas convenções, vemos que existe uma única reta passando por dois pontos distintos quaisquer de $\mathbb{P}^2(\mathbb{R})$, e além disso, quaisquer retas distintas em $\mathbb{P}^2(\mathbb{R})$ se interceptam em exatamente em um ponto.

Os pontos de $\mathbb{A}^2(\mathbb{R})$ são descritos por pares de números (x, y), onde $x, y \in \mathbb{R}$. Apresentaremos uma descrição mais precisa das direções em $\mathbb{P}^1(\mathbb{R})$. Para tanto, vamos descrevê-las através do conjunto das retas que passam pela origem $(0, 0)$, uma vez que toda reta em $\mathbb{A}^2(\mathbb{R})$ é paralela a uma reta que passa pela origem. Cada ponto $(A, B) \neq (0, 0)$ determina uma única reta que passa pela origem. Dois pontos (A, B) e $(A^{'}, B^{'})$ determinam uma mesma reta, quando existe $t \neq 0$ tal que $A = tA^{'}$ e $B = tB^{'}$. Esta relação é uma relação de equivalência em $\mathbb{A}^2(\mathbb{R})\backslash\{(0, 0)\}$. Assim, o conjunto das direções em $\mathbb{A}^2(\mathbb{R})$ é naturalmente descrito pelas classes de equivalência de pontos (A, B), com A e B não simultaneamente nulos. Denotaremos a classe de equivalência de um ponto (A, B) por $(A : B)$. Assim, $(A : B) = \{(tA, tB) \mid t \in \mathbb{R} \setminus \{0\}\}$ e

$$\begin{aligned}\mathbb{P}^1(\mathbb{R}) &= \{(A:B) \mid (A,B) \in \mathbb{A}^2(\mathbb{R}) \setminus \{(0,0)\}\} = \\ &= \{(A:B) \mid B \neq 0\} \cup \{(A:0) \mid A \in \mathbb{R} \setminus \{0\}\} = \\ &= \left\{\left(\frac{A}{B}:1\right) \mid B \neq 0\right\} \cup \{(1:0)\}\end{aligned}$$

Representando o ponto $(1:0)$ pelo símbolo ∞, podemos, então, escrever

$$\mathbb{P}^1(\mathbb{R}) = \{(x:1) \mid x \in \mathbb{R}\} \cup \{\infty\}.$$

Assim, identificamos a reta no infinito $\mathbb{P}^1(\mathbb{R})$ com a reta real $\mathbb{A}^1(\mathbb{R})$ (ou simplesmente $\mathbb{R}$) acrescida de um ponto extra, chamado de ponto no infinito (∞), através das seguintes funções:

$$\mathbb{P}^1(\mathbb{R}) = \{(A:B) \mid \text{A e B não simultaneamente nulos}\} \longleftrightarrow \mathbb{A}^1(\mathbb{R}) \cup \{\infty\}$$

$$(A:B) \qquad \longrightarrow \qquad \begin{cases} \frac{A}{B} \in \mathbb{R} & \text{se} \quad B \neq 0 \\ \infty \text{ (ponto no infinito)} & \text{se} \quad B = 0 \end{cases}$$

$$\begin{array}{ccc} (x:1) & \longleftarrow & x \in \mathbb{A}^1(\mathbb{R}) \\ (1:0) & \longleftarrow & \infty \end{array}$$

Para visualizar melhor esta identificação, mergulhamos a reta real no plano euclidiano identificando-a com a reta $y = 1$. A cada reta não horizontal que passa pela origem $(0,0)$, associamos um ponto P da reta $y = 1$. À reta $y = 0$, associamos o ponto no infinito que é a direção da reta $y = 0$. Assim, a cada reta do plano euclidiano que passa pela origem, associamos um ponto de $\mathbb{P}^1$; às retas não horizontais, associamos os pontos de $\mathbb{A}^1$, e à reta $y = 0$, o ponto no infinito. Veja a figura 3.3.

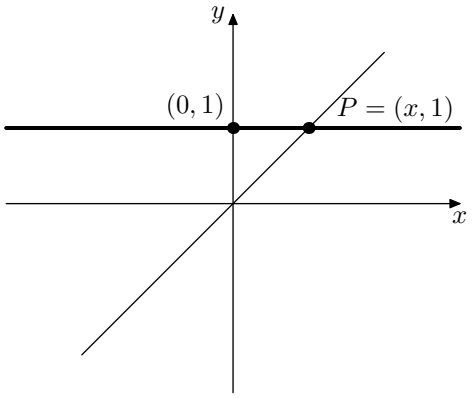


Figura 3.3

Aproveitando a geometria de $\mathbb{A}^2(\mathbb{R})$, o que fizemos foi definir o plano projetivo $\mathbb{P}^2(\mathbb{R})$ adicionando a cada reta o ponto no infinito correspondente à sua direção.

Tendo apresentado uma definição geométrica do Plano Projetivo $\mathbb{P}^2(\mathbb{R})$, vamos, agora, apresentar uma definição algébrica do mesmo e, mais adiante, mostrar que as duas definições dadas são equivalentes. Para a definição algébrica, vamos considerar o caso mais geral, onde K é um corpo arbitrário. Consideremos o conjunto de todas as ternas (x, y, z) onde x, y e z pertencem ao corpo K e não são todos nulos. Consideremos nesse conjunto a relação de equivalência $\sim$ definida por: $(x, y, z) \sim (x', y', z')$ se, e somente se, existe $t \neq 0$ tal que $x = tx'$, $y = ty'$ e $z = tz'$. Dada uma terna (x, y, z) denotamos por $(x : y : z)$ a sua classe de equivalência. O plano projetivo $\mathbb{P}^2(K)$ é, por definição, o conjunto dessas classes de equivalências, isto é,

$$\mathbb{P}^2(K) = \{(x : y : z) \mid (x, y, z) \in K^3 \setminus \{(0, 0, 0)\}\} = \frac{K^3 \setminus \{(0, 0, 0)\}}{\sim}.$$

Os números x, y e z são chamados de coordenadas homogêneas do ponto $(x : y : z)$. Dado um ponto $(x : y : z)$ do plano projetivo $\mathbb{P}^2(K)$ com $z \neq 0$, temos que

$$(x : y : z) = \left(\frac{x}{z} : \frac{y}{z} : 1\right)$$

pois

$$x = z.\frac{x}{z}, \quad y = z.\frac{y}{z} \quad \text{e} \quad z = z.1.$$

Mais geralmente, para cada inteiro $n \geq 1$, definimos o espaço projetivo $\mathbb{P}^n(K)$, de dimensão n, como sendo o espaço quociente de $K^{n+1} \setminus \{(0, 0, \ldots, 0)\}$ pela relação de equivalência $\sim$, onde duas $(n + 1)$-uplas $(x_1, x_2, \ldots, x_{n+1})$ e $(x'_1, x'_2, \ldots, x'_{n+1})$, não nulas, estão relacionadas, se existe $t \neq 0$ tal que $x_1 = tx'_1$, $x_2 = tx'_2$, ..., $x_{n+1} = tx'_{n+1}$.

Denotando por $(x_1 : \ldots : x_{n+1})$ a classe de equivalência de um ponto $(x_1, \ldots, x_{n+1}) \in K^{n+1} \setminus \{(0, 0, \ldots, 0)\}$, temos

$$\mathbb{P}^n(K) = \{(x_1 : \ldots : x_{n+1}) \mid (x_1, \ldots, x_{n+1}) \in K^{n+1} \setminus \{(0, 0, \ldots, 0)\}\}.$$

Acabamos de apresentar uma definição algébrica do Plano Projetivo $\mathbb{P}^n(K)$. Fazendo $K = \mathbb{R}$, obtemos a definição algébrica de $\mathbb{P}^2(\mathbb{R})$. O que faremos em seguida é mostrar que as definições algébrica e geométrica do Plano Projetivo são equivalentes.

Vimos, também, que a descrição geométrica do plano projetivo $\mathbb{P}^2(\mathbb{R})$ pode ser expressa na forma

$$\mathbb{P}^2(\mathbb{R}) = \mathbb{A}^2(\mathbb{R}) \cup \mathbb{P}^1(\mathbb{R}).$$

Assim, um ponto do Plano Projetivo $\mathbb{P}^2(\mathbb{R})$ é um ponto do plano afim, $\mathbb{A}^2(\mathbb{R})$, ou um ponto $(A : B) \in \mathbb{P}^1(\mathbb{R})$ que corresponde à direção da reta determinada pelo ponto (A, B).

Vamos, agora, relacionar as duas definições de plano projetivo apresentadas. Os pontos $(a : b : c)$ de $\mathbb{P}^2$ com coordenada $c \neq 0$ podem ser representados na forma $\left(\frac{a}{c} : \frac{b}{c} : 1\right)$ e estão relacionados com os pontos ordinários $\left(\frac{a}{c}, \frac{b}{c}\right)$ de $\mathbb{A}^2$; um ponto $(x, y) \in \mathbb{A}^2$ está associado a um ponto $(x : y : 1) \in \mathbb{P}^2$. E os pontos $(a : b : c)$ de $\mathbb{P}^2$ com coordenada $c = 0$? Cada um desses pontos está associado a um ponto de $\mathbb{P}^1$. A função abaixo mostra como é feita a relação entre as duas definições do plano projetivo.

$$\mathbb{P}^2 = \{(a : b : c) \mid a, b \text{ e } c \text{ não simultaneamente nulos}\} \longleftrightarrow \mathbb{A}^2 \cup \mathbb{P}^1$$

$$(a : b : c) \longrightarrow \begin{cases} \left(\frac{a}{c}, \frac{b}{c}\right) \in \mathbb{A}^2 & \text{se} \quad c \neq 0 \\ (a : b) \in \mathbb{P}^1 & \text{se} \quad c = 0 \end{cases}$$

$$(x : y : 1) \longleftarrow (x, y) \in \mathbb{A}^2$$

$$(A : B : 0) \longleftarrow (A : B) \in \mathbb{P}^1$$

Para se ter uma idéia geométrica, consideremos o plano afim $\mathbb{A}^2(\mathbb{R})$ mergulhado no espaço tridimensional como o plano π de equação $Z = 1$. Cada ponto de π determina uma única reta, não paralela ao plano $Z = 0$, que passa por este ponto e pela origem $(0, 0, 0)$. Reciprocamente, cada reta no espaço tridimensional que passa pela origem e não é paralela ao plano $Z = 0$,

determina um único ponto em π. Cada direção de uma reta contida no plano π, isto é, cada ponto no infinito de π, está associada à uma reta contida no plano $Z = 0$ que passa pela origem, e vice-versa. Assim, o plano projetivo $\mathbb{P}^2(\mathbb{R})$ é a união do plano afim $\mathbb{A}^2(\mathbb{R}) = \{(x : y : 1) \mid x, y \in \mathbb{R}\}$ (retas pela origem em $\mathbb{R}^3$, não paralelas ao plano $Z = 0$) com $\mathbb{P}^1(\mathbb{R}) = \{(x : y : 0) \mid x, y \in \mathbb{R}, x^2 + y^2 > 0\}$ (retas pela origem em $\mathbb{R}^3$, paralelas ao plano $Z = 0$).

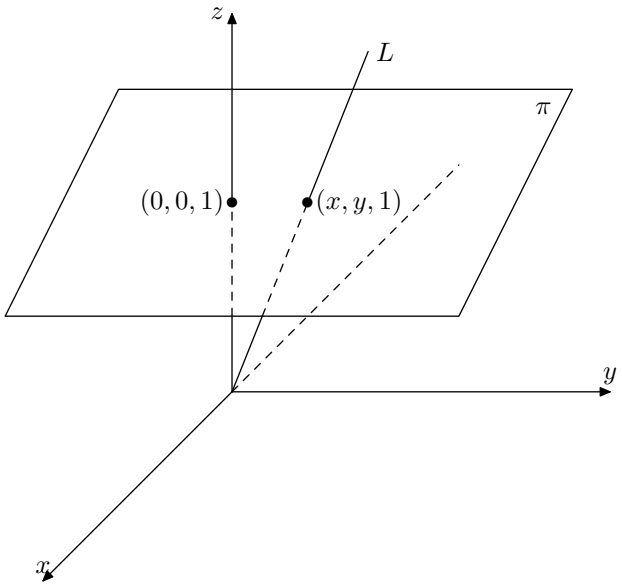


Figura 3.4

Agora, vamos apresentar a definição de retas em $\mathbb{P}^2(K)$. Mais adiante, definiremos curvas mais gerais.
Uma reta no Plano Projetivo é um conjunto de pontos $(a : b : c)$ em $\mathbb{P}^2(K)$ cujas coordenadas satisfazem uma equação da forma

$$\alpha X + \beta Y + \gamma Z = 0 \tag{3.1}$$

onde α, β e $\gamma \in K$ são constantes não simultaneamente nulas. Observe que se $(a : b : c)$ satisfaz a equação 3.1, então $(ta : tb : tc)$ também a satisfaz, para qualquer constante $t \neq 0$. Assim, para verificar se um ponto de $\mathbb{P}^2(K)$ está sobre uma reta dada, pode-se usar qualquer terna de coordenadas homogêneas para o ponto.

Cada definição do plano projetivo tem a descrição do que é uma reta. Vamos verificar que essas definições se equivalem. Por exemplo, uma reta L em $\mathbb{P}^2$ é o conjunto das soluções $(a : b : c)$ de uma equação da forma $\alpha X + \beta Y + \gamma Z = 0$. Suponhamos, primeiramente, que α e β não são simultaneamente nulos (isto significa que estamos considerando todas as retas em

$\mathbb{P}^2$, com exceção da reta $Z = 0$). Então, qualquer ponto $(a : b : c) \in L$ com $c \neq 0$ está relacionado com o ponto $(a/c, b/c)$ sobre a reta $\alpha x+\beta y+\gamma = 0$ em $\mathbb{A}^2$. O ponto $(-\beta : \alpha : 0) \in L$ está relacionado com o ponto $(-\beta : \alpha) \in \mathbb{P}^1$, que corresponde à direção da reta $-\beta y = \alpha x$. Isto está correto pois a reta $-\beta y = \alpha x$ é precisamente a reta que passa pela origem e é paralela à reta $\alpha x + \beta y + \gamma = 0$. Suponhamos, agora, que $\alpha = \beta = 0$ e, portanto $\gamma \neq 0$ (nesse caso estamos considerando a reta $Z = 0$). Essa reta está relacionada com a reta em $\mathbb{A}^2 \cup \mathbb{P}^1$ que consiste de todos os pontos no infinito. Então, as retas nas duas descrições do plano projetivo são consistentes.

Uma vez definido o Plano Projetivo, vamos, agora, estudar curvas projetivas. Já vimos um caso particular que foram as retas. Para definir curvas projetivas precisaremos usar polinômios em três variáveis, uma vez que os pontos de $\mathbb{P}^2(K)$ são representados por ternas homogêneas. Além disso, como cada ponto de $\mathbb{P}^2(K)$ pode ser representado por diferentes ternas homogêneas, só faz sentido considerar polinômios $F(X, Y, Z)$ tais que $F(a, b, c) = 0$ implica em $F(ta, tb, tc) = 0$, para todo $t \in K$. Tais polinômios são denominados polinômios homogêneos.

Definição 3.1.0.24 Dizemos que um polinômio $F(X, Y, Z)$ é homogêneo de grau d, se ele satisfaz a identidade

$$F(tX, tY, tZ) = t^d F(X, Y, Z), \forall\, t \in K.$$

A identidade que aparece na definição acima é equivalente à afirmação de que F é uma combinação linear dos monômios $X^i Y^j Z^k$ com $i + j + k = d$.

Definição 3.1.0.25 Uma curva projetiva C, no plano projetivo $\mathbb{P}^2(K)$, é o conjunto das soluções de uma equação polinomial $C : F(X, Y, Z) = 0$, onde F é um polinômio homogêneo não constante. O grau da curva C é o grau do polinômio F.

Exemplo 3.1.0.5 As curvas

$$C_1 : X^2 + Y^2 - Z^2 = 0 \quad \text{e} \quad C_2 : Y^2 Z - X^3 - XZ^2 = 0$$

são curvas projetivas, onde C_1 tem grau 2, e C_2 tem grau 3. Observe que todos os monômios que aparecem em C_1 têm grau 2, e em C_2 grau 3.

Para verificar se um ponto $P \in \mathbb{P}^2(K)$ pertence a uma curva C, podemos considerar quaisquer coordenadas homogêneas $(a : b : c)$ de P e verificar se $F(a, b, c) = 0$. Isto porque quaisquer outras coordenadas homogêneas para P são da forma (ta, tb, tc), para algum $t \neq 0$. Assim, $F(a, b, c)$ e $F(ta, tb, tc)$ são ambos iguais a zero ou ambos diferentes de zero. Isto nos dá a descrição de

uma curva projetiva, quando usamos a definição de $\mathbb{P}^2(K)$ por coordenadas homogêneas.

Vamos, agora, relacionar essa descrição de uma curva em $\mathbb{P}^2(K)$ com a sua descrição geométrica, isto é, quando olhamos $\mathbb{P}^2(K)$ como $\mathbb{A}^2 \cup \mathbb{P}^1$. Para tanto, seja $C \subset \mathbb{P}^2(K)$ uma curva dada por um polinômio homogêneo de grau d,

$$C : F(X, Y, Z) = 0.$$

Se $P = (a : b : c)$ é um ponto de $C \subset \mathbb{P}^2$ com $c \neq 0$ então, de acordo com a identificação $\mathbb{P}^2 \longleftrightarrow \mathbb{A}^2 \cup \mathbb{P}^1$ descrita anteriormente, o ponto $P \in C \subset \mathbb{P}^2$ corresponde ao ponto

$$\left(\frac{a}{c}, \frac{b}{c}\right) \in \mathbb{A}^2 \subset \mathbb{A}^2 \cup \mathbb{P}^1,$$

onde $\mathbb{A}^2$ é o plano afim usual e os pontos no infinito (pontos de $\mathbb{P}^1$) correspondem às direções em $\mathbb{A}^2$. Como F é homogêneo de grau d, e $F(a, b, c) = 0$, temos que

$$F\left(\frac{a}{c}, \frac{b}{c}, 1\right) = F\left(\frac{1}{c}(a, b, c)\right) = \frac{1}{c^d} F(a, b, c) = 0.$$

Em outras palavras, se definirmos um novo polinômio $f(x, y)$ por $f(x, y) = F(x, y, 1)$, obteremos uma função

$$\{(a : b : c) \in C \quad | \quad c \neq 0\} \quad \longrightarrow \quad \{(x, y) \in \mathbb{A}^2 \quad | \quad f(x, y) = 0\}.$$

$$(a : b : c) \longmapsto \left(\frac{a}{c}, \frac{b}{c}\right)$$

Essa função é bijetiva, visto que se $(r, s) \in \mathbb{A}^2$ satisfaz a equação $f(x, y) = 0$, então $(r : s : 1) \in C$. Chamamos a curva $f(x, y) = 0$ em $\mathbb{A}^2$ de a *parte afim* da curva projetiva C.

Falta olhar para os pontos $(a : b : c) \in C$ com $c = 0$ e descrevê-los geometricamente em termos da parte afim de C. Os pontos $(a : b : 0) \in C$ satisfazem a equação $F(X, Y, 0) = 0$ e são enviados aos pontos no infinito $(a : b) \in \mathbb{P}^1$ em $\mathbb{A}^2 \cup \mathbb{P}^1$. Afirmamos que esses pontos, que como já vimos, são direções em $\mathbb{A}^2$, correspondem às assíntotas da curva afim $f(x, y) = 0$. Em outras palavras, uma curva afim $f(x, y) = 0$ tem, de alguma maneira, alguns pontos faltando, pontos esses que estão no infinito e são as direções limite quando se caminha ao longo da curva para o infinito. Vamos ilustrar com dois exemplos a relação dos pontos de uma curva projetiva com terceira coordenada zero com os pontos no infinito em $\mathbb{A}^2 \cup \mathbb{P}^1$. Consideremos

primeiramente uma reta $L : \alpha X + \beta Y + \gamma Z = 0$, digamos com $\alpha \neq 0$. A parte afim de L é a reta $L_0 : \alpha x + \beta y + \gamma = 0$ em $\mathbb{A}^2$. Os pontos com coordenada $Z = 0$ correspondem aos pontos no infinito sobre L. Existe somente um tal ponto, a saber o ponto $(-\beta : \alpha : 0) \in L$, que corresponde ao ponto $(-\beta : \alpha) \in \mathbb{P}^1$, que por sua vez corresponde à direção da reta $-\beta y = \alpha x$ em $\mathbb{A}^2$. Essa direção é exatamente a direção da reta L_0. Assim L consiste da reta afim L_0 juntamente com o ponto no infinito que corresponde à direção de L_0.

Vejamos, agora, um outro exemplo. Consideremos a curva projetiva

$$C : X^2 - Y^2 - Z^2 = 0.$$

Existem dois pontos em C com $Z = 0$, a saber $(1 : 1 : 0)$ e $(1 : -1 : 0)$. Esses dois pontos correspondem, respectivamente, aos dois pontos no infinito $(1 : 1), (1 : -1) \in \mathbb{P}^1$, ou equivalentemente, às direções das retas $y = x$ e $y = -x$ em $\mathbb{A}^2$. A parte afim da curva C que obtemos fazendo $Z = 1$ na equação acima é a hipérbole

$$C_0 : x^2 - y^2 - 1 = 0.$$

Seja L a reta tangente à curva C_0 num ponto (r, s). Quando o ponto (r, s) se desloca sobre a curva, por exemplo para $\mid r \mid$ tendendo a $+\infty$, a direção da reta tangente L se aproxima da direção de uma das retas $y = x$ e $y = -x$, que são as assíntotas da curva C_0.

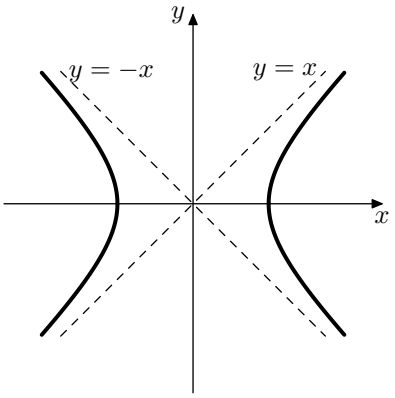


Figura 3.5

A discussão acima mostra que se começamos com uma curva projetiva $C : F(X, Y, Z) = 0$, então podemos escrever C como a união de sua parte afim C_0 e seus pontos no infinito. Aqui C_0 é a curva afim dada pela equação

$$C_0 : f(x, y) = F(x, y, 1) = 0,$$

e os pontos no infinito de C são os pontos com $Z = 0$, que correspondem às direções limites das retas tangentes à C_0. O processo de trocar o polinômio homogêneo $F(X, Y, Z)$ pelo polinômio $f(x, y) = F(x, y, 1)$ chama-se desomogenização (com relação à variável Z). Veremos agora o processo inverso.

Comecemos com uma curva afim C_0 dada por uma equação $f(x, y) = 0$. Queremos achar uma curva projetiva C cuja parte afim é C_0 ou, equivalentemente, queremos achar um polinômio homogêneo $F(X, Y, Z)$ tal que $F(x, y, 1) = f(x, y)$. Isto é fácil de fazer, embora tenhamos de tomar cuidado para não incluir também a reta no infinito L_∞ em nossa curva. Definimos a homogenização do polinômio $f(x, y) = \sum a_{ij}x^i y^j$ como sendo

$$F(X, Y, Z) = \sum_{i,j} a_{ij}X^i Y^j Z^{d-i-j}, \qquad \text{onde } d = gr(f).$$

Desta definição, segue que $F(x, y, 1) = f(x, y)$. Além disso, nossa escolha de d assegura que $F(X, Y, 0)$ não é identicamente nulo e, portanto a curva definida por $F(X, Y, Z) = 0$ não contém a reta no infinito L_∞. Se escolhêssemos $d > gr(f)$, todo monômio $X^i Y^j Z^{d-i-j}$ em F teria um fator Z e portanto $F(X, Y, 0)$ seria identicamente nulo. Vemos, então, que, usando homogenização e desomogenização, obtemos uma correspondência um a um entre curvas afins e curvas projetivas que não contêm a reta no infinito.

Devemos também mencionar que não há nada em especial com a variável Z. Quando consideramos uma curva em $\mathbb{P}^2$ dada por um polinômio podemos desomogenizá-lo com relação a uma das outras duas variáveis. No caso visto acima, poderíamos ter desomogenizado o polinômio $F(X, Y, Z)$ com relação a Y e obteríamos a curva $F(x, 1, z) = 0$ no plano afim xz. Às vezes, é conveniente fazer isto, se estivermos interessados em algum dos pontos no infinito de uma curva projetiva C. Em essência, o que estamos fazendo é tomar uma reta diferente, a reta $Y = 0$, para ser a reta no infinito L_∞. Vamos apresentar um exemplo para tornar isto claro. Suponhamos que queremos estudar a curva

$$C : Y^2Z - X^3 - Z^3 = 0 \text{ e o ponto } P = (0 : 1 : 0) \in C.$$

Se nós desomogenizarmos com relação a Z, então o ponto P tornar-se-á um ponto no infinito sobre a curva afim $y^2 - x^3 - 1 = 0$. Em vez disso, desomogeneizamos com relação a Y, o que significa fazer $Y = 1$. Obtemos assim a curva afim

$$C : z - x^3 - z^3 = 0 \text{ e o ponto } P \text{ torna-se o ponto } (x, z) = (0, 0).$$

Em geral, considerando diferentes retas para ser a reta no infinito, podemos decompor uma curva projetiva C em várias partes afins sobrepostas, e então

essas partes afins podem ser "coladas"de modo formar a curva projetiva inteira.

Ao estudar interseções de duas curvas planas, é importante olhar para as retas tangentes às curvas nos pontos de interseção. Sabe-se do Cálculo Diferencial que dados uma curva plana afim C de equação $f(x, y) = 0$ e um ponto $P = (r, s) \in C$, a reta tangente à curva C em P é dada pela equação

$$\frac{\partial f}{\partial x}(r, s)(x - r) + \frac{\partial f}{\partial y}(r, s)(y - s) = 0.$$

No entanto, se ambas as derivadas parciais forem iguais a zero, teremos um problema. Isto acontece para cada uma das curvas

$$C_1 : y^2 = x^3 + x^2 \qquad \text{e} \qquad C_2 : y^2 = x^3$$

no ponto $P = (0, 0)$. Veja os gráficos de C_1 e C_2 nas figuras abaixo.

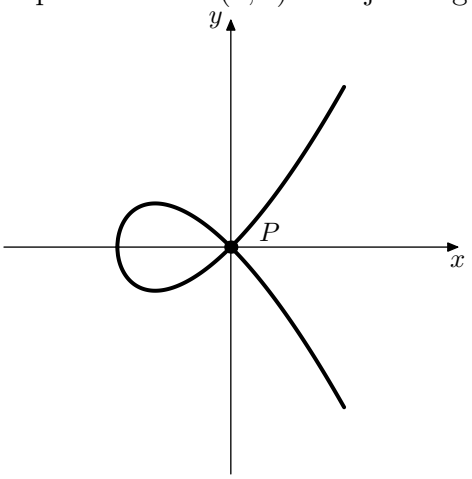


Figura 3.6

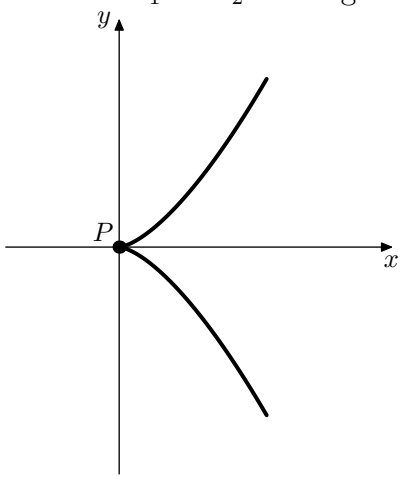


Figura 3.7

A curva C_1 intercepta a si própria em P, tendo, portanto, duas direções distintas neste ponto. A curva C_2, por outro lado, tem uma cúspide em P, em outras palavras, tem um ponto anguloso em P.

Definição 3.1.0.26 Dizemos que P é um ponto singular de uma curva $C : f(x, y) = 0$, se

$$\frac{\partial f}{\partial x}(P) = \frac{\partial f}{\partial y}(P) = 0.$$

Dizemos, também, que C é uma curva não-singular (ou curva suave), se todo ponto de C é não-singular. Se P é um ponto não-singular de C, então definimos a reta tangente a C em P como sendo a reta descrita acima.

Para uma curva projetiva C descrita por um polinômio homogêneo $F(X,Y,Z) = 0$ apresentamos definições análogas. Mais precisamente, se $P = (a : b : c)$ é um ponto de C com $c \neq 0$, então consideramos a parte afim de C e verificamos se o ponto $P_0 = \left(\frac{a}{c}, \frac{b}{c}\right)$ é ou não um ponto singular da curva afim $C_0 : F(x,y,1) = 0$. Se $c = 0$, então podemos desomogeneizar de alguma outra maneira. Por exemplo, se $a \neq 0$, verificamos se o ponto $P_0 = \left(\frac{b}{a}, \frac{c}{a}\right)$ é ou não um ponto singular da curva afim $C_0 : F(1,y,z) = 0$. Resumindo, um ponto P em uma curva projetiva C é singular, se ele é singular para alguma parte afim C_0 de C. Dizemos, então, que uma curva projetiva C é não singular ou suave, se todos os seus pontos, incluindo os pontos no infinito são não-singulares. Se P é um ponto não-singular de C, definimos a reta tangente a C em P por desomogenização, achando a reta tangente à parte afim de C em P, e então homogeneizando a equação da reta tangente para obter uma reta em $\mathbb{P}^2$. Podemos provar que um ponto P em uma curva projetiva $C : F(X,Y,Z) = 0$ é singular se, e somente se,

$$\frac{\partial F}{\partial X}(P) = \frac{\partial F}{\partial Y}(P) = \frac{\partial F}{\partial Z}(P) = 0.$$

Para concluir, observamos que quando consideramos um corpo qualquer K, $K \neq \mathbb{R}$, pode não ter sentido falar em derivada. O que fazemos é definir derivada de um polinômio $\sum_{k=0}^{n} a_k x^k$ algebricamente como sendo $\sum_{k=1}^{n} k a_k x^{k-1}$.

Vamos terminar esta seção comentando a respeito da geometria das curvas projetivas quando K não for o corpo dos números reais. Em outras palavras, como ficam nossas intuições geométricas quando $K = \mathbb{Z}_p$, por exemplo? Nesse caso, os polinômios têm coeficientes em $\mathbb{Z}_p$ e suas soluções também estão em $\mathbb{Z}_p$. Mas o que dizer a respeito de pontos, curvas e direções em $\mathbb{A}^2(K)$ quando $K = \mathbb{Z}_p$? Na verdade, podemos continuar pensando no plano euclidiano que a maioria das nossas intuições geométricas ainda serão verdadeiras quando trocarmos as coordenadas para $\mathbb{Z}_p$. Além disso, observamos que os planos afim e projetivo e as curvas afim e projetivas são definidas algebricamente em termos de pares ordenados (r,s) ou ternas homogêneas $(a : b : c)$ sem qualquer referência à geometria. Assim podemos provar resultados trabalhando algebricamente usando coordenadas, sem nos preocupar com intuições geométricas. Resumindo o que foi dito, devemos pensar geometricamente e provar algebricamente.

3.2 Interseções de Curvas Projetivas

Vimos que duas retas quaisquer no plano projetivo $\mathbb{P}^2$, isto é, curvas cujas equações são dadas por polinômios de grau 1, se interceptam em exatamente um ponto. O teorema de Bezout, que será enunciado no final desta seção, nos diz que o número dos pontos de interseção de duas curvas projetivas é sempre igual ao produto dos graus dessas curvas, desde que interpretemos esta afirmação adequadamente. Vamos apresentar alguns exemplos de interseções de curvas de graus maiores que 1 e algumas considerações adicionais com o objetivo de entender o enunciado do teorema de Bezout.

Exemplo 3.2.0.6 $C_1 : x + y + 1 = 0$ e $C_2 : x^2 + y^2 = 1$.
Nesse caso temos $C_1 \cap C_2 = \{(-1, 0), (0, -1)\}$, o que é facilmente verificado.

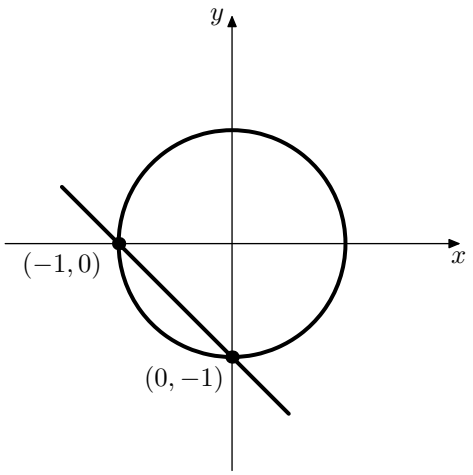


Figura 3.8

Exemplo 3.2.0.7 $C_1 : x + y = 0$ e $C_2 : x^2 + y^2 = 1$.
Nesse caso temos também dois pontos de intersecção. Porém, observamos que embora as duas curvas sejam racionais os dois pontos de intersecção $(\frac{1}{2}\sqrt{2}, -\frac{1}{2}\sqrt{2})$ e $(-\frac{1}{2}\sqrt{2}, \frac{1}{2}\sqrt{2})$ tem coordenadas em $\mathbb{R} \setminus \mathbb{Q}$.

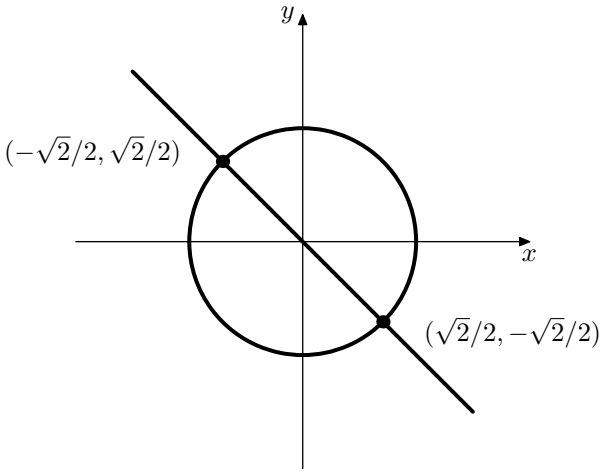


Figura 3.9

Exemplo 3.2.0.8 $C_1 : x + y + 2 = 0$ e $C_2 : x^2 + y^2 = 1$.
Essas duas curvas não se interceptam no plano euclidiano usual $\mathbb{R}^2$, mas se nós permitirmos coordenadas complexas, então encontraremos também dois pontos de intersecção $\left(-1 + \frac{\sqrt{2}}{2}i, -1 - \frac{\sqrt{2}}{2}i\right)$ e $\left(-1 - \frac{\sqrt{2}}{2}i, -1 + \frac{\sqrt{2}}{2}i\right)$.

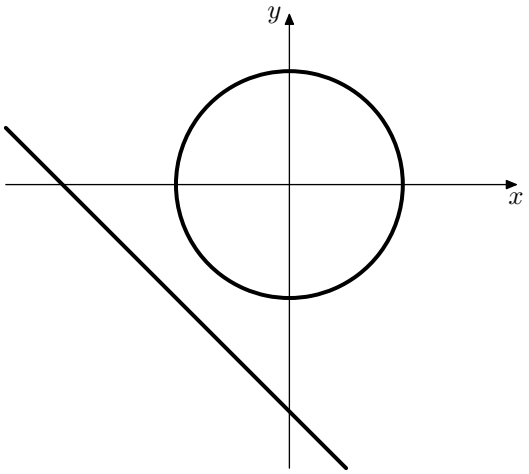


Figura 3.10

Exemplo 3.2.0.9 $C_1 : x + 1 = 0$ e $C_2 : x^2 - y = 0$.
No plano euclidiano usual, estas curvas se interceptam em um único ponto. Mas lembremos que mesmo para duas retas podemos precisar olhar para os pontos no infinito de $\mathbb{P}^2$. Neste exemplo, a reta C_1 é uma reta vertical e as retas tangentes à parábola C_2 se aproximam da direção vertical. Assim,

geometricamente, C_1 e C_2 deveriam ter um ponto no infinito em comum correspondendo à direção vertical. Para verificar isto algebricamente, primeiro, homogeneizamos as equações de C_1 e C_2 para obter as curvas $\overline{C}_1$ e $\overline{C}_2$ em $\mathbb{P}^2$:

$$\overline{C}_1 : X + Z = 0 \quad \text{e} \quad \overline{C}_2 : X^2 - YZ = 0.$$

Fazendo $X = -Z$ na equação de C_2, obtemos os dois pontos de interseção $(-1 : 1 : 1)$ e $(0 : 1 : 0)$. Assim, trabalhando com curvas projetivas, encontramos os dois pontos esperados.

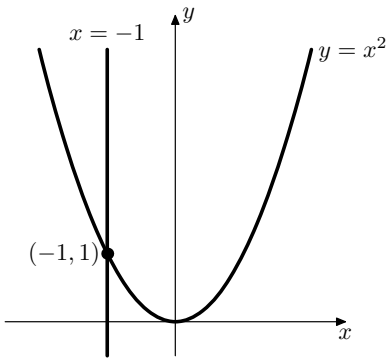


Figura 3.11

Consideremos uma situação onde um outro tipo de problema ocorre.

Exemplo 3.2.0.10 $C_1 : x + y = 2$ e $C_2 : x^2 + y^2 = 2$.
O conjunto $C_1 \cap C_2$ consiste de um único ponto $(1, 1)$, e mesmo se nós considerarmos as curvas projetivas

$$\overline{C}_1 : X + Y = 2Z \quad \text{e} \quad \overline{C}_2 : X^2 + Y^2 = 2Z^2$$

ainda achamos um único ponto de interseção $(1 : 1 : 1)$.

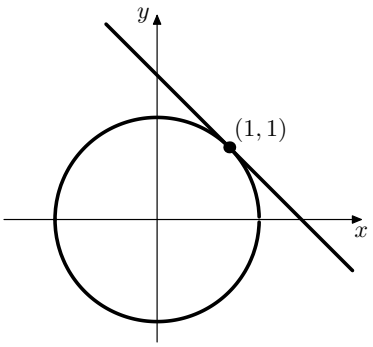


Figura 3.12

Geometricamente, vemos, com clareza, qual é o problema; é que a reta C_1 é tangente ao círculo C_2 no ponto $(1,1)$ e, portanto, em algum sentido, o ponto deveria ser contado duas vezes. Também podemos ver isto algebricamente. Se substituirmos a equação $y = x - 2$ na equação C_2 e simplificarmos a equação resultante, obteremos a equação $2x^2 - 4x + 2 = 0$, ou equivalentemente $2(x-1)^2 = 0$. Assim obtemos uma equação quadrática na variável x, e normalmente esperaríamos achar duas raízes distintas, mas neste caso encontramos uma raiz repetida. Isto faz sentido, uma vez que, mesmo para um polinômio de grau d em uma variável só podemos dizer que ele tem d raízes complexas se contarmos raízes repetidas de acordo com a suas multiplicidades.

O problema da multiplicidade pode também ocorrer, se uma das curvas for singular em P, mesmo que as duas curvas não tenham a mesma direção tangente. Por exemplo, consideremos a interseção da reta e da curva de grau três dadas a seguir.

Exemplo 3.2.0.11 $C_1 : x - y = 0$ e $C_2 : x^3 - y^2 = 0$.

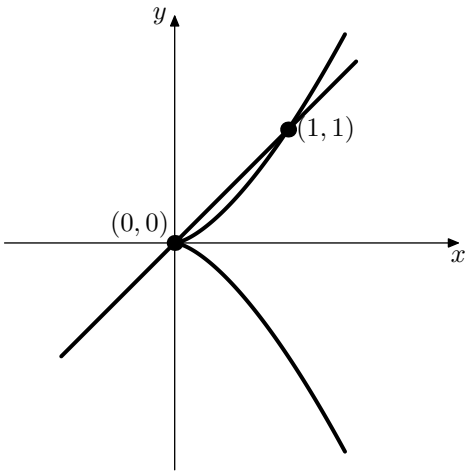


Figura 3.13

Nossa intuição nos diz que $C_1 \cap C_2$ deveria consistir de três pontos. Substituindo $y = x$ na equação de C_2 obtemos $x^3 - x^2 = 0$. Isto é uma cúbica na variável x, mas tem somente duas raízes distintas, a saber $x = 0$ e $x = 1$. Assim, $C_1 \cap C_2$ contém apenas os dois pontos $(0,0)$ e $(1,1)$, mas o ponto $(0,0)$ deve ser contado duas vezes, o que dá os três pontos esperados.

Finalmente, consideremos a interseção da reta e da cônica abaixo.

Exemplo 3.2.0.12 $C_1 : x+y+1=0$ e $C_2 : 2x^2+xy-y^2+4x+y+2=0$. Quando substituímos $y=-x-1$ na equação de C_2 vemos que tudo se cancela obtendo assim $0=0$. Isto acontece porque a equação de C_2 se fatora como

$$2x^2+xy-y^2+4x+y+2=(x+y+1)(2x-y+2),$$

portanto todo ponto de C_1 está em C_2. Observe que C_2 é a reunião de duas curvas a saber C_1 e a reta $2x-y+2=0$.

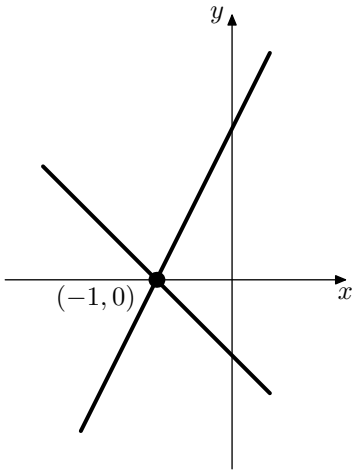


Figura 3.14

Antes de prosseguir, vamos apresentar algumas definições a respeito de polinômios de uma e duas variáveis.

Definição 3.2.0.27 Seja K um corpo. Dizemos que um polinômio não constante $p \in K[x]$ é irredutível em $K[x]$ ou irredutível sobre K, se ele não pode ser escrito como um produto de dois polinômios, não constantes, de graus menores, com coeficientes em $K[x]$. Um polinômio $g \in K[x]$ não constante e não irredutível chama-se redutível ou composto.

Exemplo 3.2.0.13 O polinômio x^2+1 é irredutível sobre $\mathbb{R}[x]$, mas não é irredutível em $\mathbb{C}[x]$, pois pode ser fatorado como $x^2+1=(x-i)\cdot(x+i)$, onde i é tal que $i^2=-1$.

Todo polinômio não constante $p \in K[x]$ pode ser escrito como produto de fatores irredutíveis em $K[x]$.

Valem resultados análogos para polinômios em duas variáveis.

Definição 3.2.0.28 Seja K um corpo. Dizemos que um polinômio $p \in K[x, y]$ é irredutível em $K[x, y]$ ou irredutível sobre K se ele não pode ser escrito como um produto de dois polinômios de graus menores, com coeficientes em $K[x, y]$. Um polinômio $g \in K[x, y]$ não constante e não irredutível chama-se redutível ou composto.

Exemplo 3.2.0.14 Dependendo do corpo K, um polinômio em $K[x, y]$ pode ser irredutível ou não. O polinômio $x^2 - 2y^2 = 0$ é irredutível em $\mathbb{Q}[x]$, mas não é irredutível em $\mathbb{R}[x]$, pois pode ser fatorado como $(x - \sqrt{2}y).(x + \sqrt{2}y)$.

Dado um polinômio não constante qualquer $p \in K[x, y]$, podemos escrevê-lo como produto de fatores irredutíveis em $K[x, y]$. Em geral, se uma curva C é dada por uma equação $f(x, y) = 0$, então fatoramos f como um produto de polinômios irredutíveis

$$f(x, y) = p_1(x, y) \cdot p_2(x, y) \cdot \ldots \cdot p_n(x, y).$$

Lembremos que todo polinômio em $K[x, y]$, o anel dos polinômios nas variáveis x e y com coeficientes em K, tem uma única fatoração em um produto desta forma. Dizemos que as curvas

$$p_1(x, y) = 0, \quad p_2(x, y) = 0, \quad \ldots \quad , p_n(x, y) = 0.$$

são as componentes irredutíveis da curva C. Dizemos que uma curva C é irredutível se, e somente se, ela tiver apenas uma componente irredutível, ou equivalentemente, se $f(x, y)$ for um polinômio irredutível. Se C_1 e C_2 são duas curvas, dizemos que C_1 e C_2 não têm componentes em comum se suas componentes irredutíveis são distintas. Sabe-se que $C_1 \cap C_2$ consiste de um conjunto finito de pontos se, e só se, C_1 e C_2 não têm componentes em comum. Finalmente, ao lidarmos com curvas projetivas C_1 e C_2, consideramos as mesmas definições usando fatorações em produtos de polinômios homogêneos irredutíveis em $K[X, Y, Z]$.

Consideremos agora o caso geral de curvas projetivas C_1 e C_2, sobre o corpo $\mathbb{C}$ dos números complexos, que supomos não ter componentes em comum. A interseção $C_1 \cap C_2$ é portanto um conjunto finito de pontos com coordenadas complexas. A cada ponto $P \in \mathbb{P}^2$ definimos uma *multiplicidade* ou *índice de interseção* $I(C_1 \cap C_2, P)$. Este índice é um inteiro não negativo que mede o quanto C_1 e C_2 são tangentes uma a outra, ou não suaves em P. Não daremos uma definição formal do índice de interseção, mas podemos ter uma boa idéia do que ele representa através das seguintes propriedades:

i) Se $P \notin C_1 \cap C_2$, então $I(C_1 \cap C_2, P) = 0$.

ii) Se $P \in C_1 \cap C_2$, se P é um ponto não-singular de C_1 e de C_2, e se C_1 e C_2 têm direções tangentes diferentes em P, então $I(C_1 \cap C_2, P) = 1$. (Dizemos neste caso que C_1 e C_2 se interceptam transversalmente em P).

iii) Se $P \in C_1 \cap C_2$ e se C_1 e C_2 não se interceptam transversalmente em P, então $I(C_1 \cap C_2, P) \geq 2$.

Feitas estas considerações, vamos agora enunciar o teorema de Bezout, onde consideramos K um corpo algebricamente fechado, por exemplo $\mathbb{C}$. Um corpo K é algebricamente fechado, se todo polinômio $F \subset K[x]$, de grau $n \geq 1$, pode ser fatorado em $K[x]$ como um produto de fatores lineares.

Teorema 3.2.0.1 (Teorema de Bezout) Sejam C_1 e C_2 curvas projetivas, definidas num corpo K algebricamente fechado , sem componentes em comum. Então,

$$\sum_{P \in C_1 \cap C_2} I(C_1 \cap C_2, P) = (grC_1)(grC_2),$$

onde a soma está definida sobre todos os pontos de $C_1 \cap C_2$. Em particular, se C_1 e C_2 são curvas suaves com interseções transversais apenas, então $\#(C_1 \cap C_2) = (grC_1)(grC_2)$; e em todos os casos vale a desigualdade

$$\#(C_1 \cap C_2) \leq (grC_1)(grC_2).$$

3.3 As curvas cúbicas e a lei de grupo

Nesta seção, vamos restringir nosso estudo de curvas algébricas às curvas cúbicas. Queremos definir uma estrutura de grupo no conjunto dos pontos de uma curva cúbica dada.

A equação geral de uma cúbica nas variáveis x e y é da forma

$$ax^3 + bx^2y + cxy^2 + dy^3 + ex^2 + fxy + gy^2 + hx + iy + j = 0,$$

ou na forma homogênea

$$aX^3+bX^2Y+cXY^2+dY^3+eX^2Z+fXYZ+gY^2Z+hXZ^2+iYZ^2+jZ^3 = 0,$$

nas variáveis X, Y e Z. Lembremos que estas equações podem também representar uma cúbica degenerada, como a união de três retas ou a união de uma reta e uma cônica.

Sejam C uma cúbica e L uma reta não contida em C. Como uma reta tem grau 1, e uma cúbica grau 3, segue, do Teorema de Bezout, que teremos três pontos de interseção entre elas desde que, é claro, consideremos curvas projetivas com coordenadas em um corpo algebricamente fechado, por exemplo $\mathbb{C}$, e contemos multiplicidades. Considerando isto, dados dois pontos P e Q sobre uma cúbica C, podemos obter um terceiro ponto sobre ela traçando a reta L que passa por P e Q e encontrando assim o ponto de interseção entre C e L. Pelo Teorema de Bezout tal ponto sempre existe e será denotado por $P * Q$. Se tivermos apenas um ponto P em C, traçamos a reta L, tangente à C em P. Nesse caso, a reta tangente encontra a cúbica duas vezes em P, e encontramos o terceiro ponto $P * P$ em $L \cap C$. Veja as figuras.

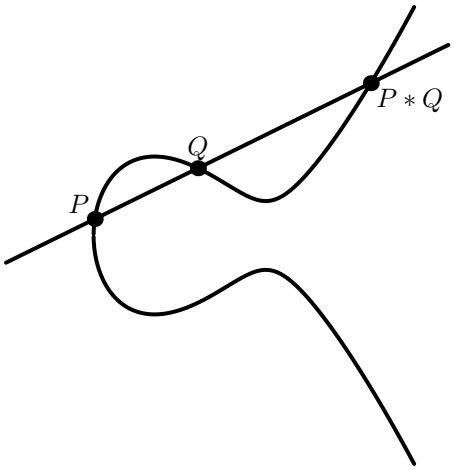


Figura 3.15

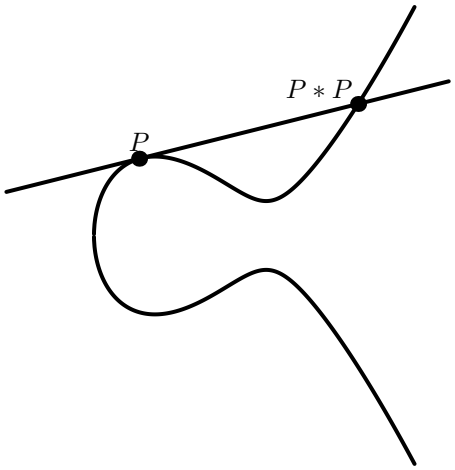


Figura 3.16

Para assegurar que o ponto $P*Q$ esteja bem definido, dois tipos de cúbicas devem ser excluídos: as cúbicas redutíveis e as cúbicas singulares. Se uma cúbica C não for irredutível, e L for uma reta contida em C, então $P*Q$ não será único, quando P e Q estiverem em L. Por outro lado, se uma curva C tiver um ponto singular P, então a reta tangente à C em P não estará bem definida, e $P * P$ não será unicamente determinado.

Lembremos que estamos considerando curvas projetivas definidas em $\mathbb{P}^2(K)$ onde K é um corpo algebricamente fechado. Nos casos em que K não é um corpo algebricamente fechado, já vimos que uma equação que define uma curva plana pode mesmo representar o conjunto vazio.

A proposição a seguir nos mostra que, se considerarmos, no entanto, uma cúbica não singular e irredutível, definida em um corpo K qualquer (não necessariamente algebricamente fechado), dados dois pontos P e Q sobre esta cúbica, o terceiro ponto de interseção, $P * Q$, estará bem definido.

Proposição 3.3.0.25 Sejam C uma cúbica irredutível e não singular e L

uma reta definida sobre um corpo K. Se a cúbica C tem dois pontos de interseção (contando suas multiplicidades) com a reta L, então C tem três pontos de interseção (contando suas multiplicidades) com a reta L.

Prova: Seja a reta $L : aX + bY + cZ = 0$ onde, por simetria, supomos que $c \neq 0$. Os pontos de interseção de C e L são as raízes do polinômio

$$q(X, Y) = p\left(X, Y, -\frac{aX + bY}{c}\right)$$

onde p é o polinômio homogêneo de grau 3 que define a curva.

Sejam $P_1 = (a_1, b_1, c_1)$ e $P_2 = (a_2, b_2, c_2)$ (podendo ser iguais), pontos da interseção de C com L. Como $q(a_1, b_1) = q(a_2, b_2) = 0$, temos que

$$q(X, Y) = v(X, Y)\prod_{i=1}^{2}(b_i X - a_i Y)$$

onde $v(X, Y)$ é um polinômio homogêneo de grau 1.
O terceiro ponto de interseção de C com L é dado por

$$P_3 = \left(a_3, b_3, -\frac{aa_3 + bb_3}{c}\right)$$

onde (a_3, b_3) é a única raiz de $v(X, Y)$. □

Podemos, agora, definir a lei de composição, que vamos chamar lei de composição secante-tangente.

1. Se P e $Q \in C(K)$ e se $P \neq Q$, então definimos $P * Q$ como sendo o terceiro ponto de interseção, contendo multiplicidades, da reta L que passa por P e Q, com a curva C.

2. Se $P \in C(K)$, então definimos $P * P$ como sendo o terceiro ponto de interseção, contendo multiplicidades, da reta L, tangente à curva C em P, com a curva C.

O que vimos até agora nos motiva a dar a seguinte definição.

Definição 3.3.0.29 Uma curva elítica $C(K)$ em $\mathbb{P}^2(K)$ é uma curva cúbica não singular e irredutível sobre K.

As curvas elíticas são precisamente as curvas cúbicas para as quais a operação $*$, que a cada par de pontos P e Q associa o ponto $P * Q$, está bem definida para todos os pares de pontos da curva. Notemos, porém, que a regra da secante-tangente não define uma estrutura de grupo em $C(K)$. Na verdade, não temos nem o elemento neutro, isto é, não existe um único ponto $\mathcal{O} \in C(K)$ tal que $P * \mathcal{O} = P$ para todo $P \in C(K)$. Se existisse um tal elemento neutro $\mathcal{O}$ para a operação $*$, então para qualquer $P \in C(K)$ teríamos $P * \mathcal{O} = P$. Assim, a reta passando por P e $\mathcal{O}$ seria a reta tangente à curva em P; portanto teríamos $P * P = \mathcal{O}$, para todo $P \in C(K)$, o que certamente não ocorre. No entanto, podemos definir uma outra operação a partir da operação $*$ que tornará $C(K)$ um grupo. Antes, porém, faremos algumas observações a respeito de relações entre curvas e equações lineares e, em seguida, apresentaremos uma propriedade geométrica elementar das curvas cúbicas.

Seja $\mathcal{C}^3$ o conjunto de todas as cúbicas em $\mathbb{P}^2$. Então, cada curva $C \in \mathcal{C}^3$ é dada por uma equação do tipo

$$aX^3+bX^2Y+cXY^2+dY^3+eX^2Z+fXYZ+gY^2Z+hXZ^2+iYZ^2+jZ^3 = 0.$$

Assim, C é determinada pelos coeficientes, $a, b, c, d, e, f, g, h, i, j$. Se multiplicarmos a equação de C por uma constante não nula, obteremos a mesma curva; então, na realidade, C é determinada pela 10-upla homogênea $(a : b : c : d : e : f : g : h : i : j)$. Reciprocamente, se duas 10-uplas representam a mesma curva, então elas diferem por uma constante multiplicativa. Em outras palavras, o conjunto $\mathcal{C}^3$ das curvas em $\mathbb{P}^2(K)$ é, de uma maneira natural, isomorfo ao espaço projetivo $\mathbb{P}^9(K)$. Consideremos o conjunto de todas as cúbicas que passam por um ponto dado $P \in \mathbb{P}^2(K)$. Tal conjunto é isomorfo a um subconjunto de $\mathbb{P}^9$. Se (X_0, Y_0, Z_0) são as coordenadas homogêneas de P, isto é, $P = (X_0 : Y_0 : Z_0)$, o conjunto das cúbicas que passam por P é o conjunto das 10-uplas $(a : b : c : d : e : f : g : h : i : j)$ que satisfazem a equação homogênea linear

$$X_0^3a + X_0^2Y_0b + X_0Y_0^2c + Y_0^3d + X_0^2Z_0e + X_0Y_0Z_0f + Y_0^2Z_0g + X_0Z_0^2h + Y_0Z_0^2i + Z_0^3j = 0,$$

nas variáveis $a, b, c, d, e, f, g, h, i, j$. Em outras palavras, dado um ponto $P \in \mathbb{P}^2(K)$, o conjunto das cúbicas $C \in \mathcal{C}^3$ que contêm P corresponde ao conjunto das soluções de uma equação linear homogênea em $\mathbb{P}^9(K)$. Do mesmo modo, dados dois pontos P e $Q \in \mathbb{P}^2$, o conjunto das cúbicas que contêm P e Q corresponde ao conjunto das soluções simultâneas de duas equações lineares em $\mathbb{P}^9$, sendo uma delas definida por P e a outra por Q. Continuando este raciocínio, vemos que existe uma bijeção entre o conjunto

das curvas que passam por n pontos, $P_1, P_2, ..., P_n$, em $\mathbb{P}^2$ e o conjunto das soluções simultâneas de um certo sistema de n equações lineares homogêneas em $\mathbb{P}^9$.

Pelo Teorema de Bezout, duas cúbicas se interceptam em nove pontos. A propriedade geométrica mencionada é a seguinte:

Propriedade: "Sejam C_1 e C_2 duas cúbicas que se interceptam em 9 pontos distintos. Se C é uma curva que passa por 8 dos 9 pontos de interseção, então C passa pelo nono ponto de interseção".

Prova: Sejam $P_1, P_2, ..., P_9$ os nove pontos distintos de interseção das cúbicas C_1 e C_2 em $\mathbb{P}^2$. Suponhamos que a cúbica C passe por oito dos nove desses pontos, a saber $P_1, P_2, ..., P_8$. Vamos mostrar que C também passa pelo ponto P_9. Consideremos que C_1 e C_2 são dadas pelas equações $C_1 : F_1(X, Y, Z) = 0$ e $C_2 : F_2(X, Y, Z) = 0$. O conjunto das curvas que passam pelos 8 primeiros pontos $P_1, P_2, ..., P_8$ corresponde ao conjunto das soluções simultâneas de 8 equações lineares homogêneas em 10 variáveis. O conjunto das soluções deste sistema pode ter dimensão maior do que 2, mas como os 8 pontos são dois a dois distintos, a dimensão do conjunto das soluções é igual a 2. Sendo assim, sejam v_1 e v_2 duas soluções independentes. Então, toda solução do sistema é da forma $\lambda_1 v_1 + \lambda_2 v_2$ para constantes λ_1 e λ_2.

Como C_1 e C_2 são duas cúbicas que passam pelos 8 pontos $P_1, P_2, ..., P_8$, os coeficientes de sua equações F_1 e F_2 formam duas 10-uplas que sãos independentes e que são soluções do sistema linear de 8 equações lineares; portanto elas geram o conjunto de todas as soluções. Como a cúbica C em $\mathbb{P}^2$ contém os 8 pontos $P_1, P_2, ..., P_8$, a equação para C será da forma

$$C : \lambda_1 F_1(X, Y, Z) + \lambda_2 F_2(X, Y, Z) = 0,$$

para constantes λ_1 e λ_2. Como o nono ponto P_9 está em ambas as curvas C_1 e C_2, temos que $F_1(P_9) = F_2(P_9) = 0$. Segue da equação acima que C também contém o nono ponto.□

Vamos, agora, utilizando a lei de composição secante-tangente, definir uma estrutura de grupo no conjunto dos pontos de uma dada cúbica C que suporemos irredutível e sem pontos singulares. Seja, então, $\mathcal{O}$ um ponto qualquer sobre C. Consideremos a operação $+$ que a cada par (P, Q) de pontos de C associa o ponto $P + Q$ sobre C definido por $P + Q = \mathcal{O} * (P * Q)$. Assim, dados dois P e Q, encontramos o terceiro ponto $P * Q$ traçando a

reta L_1 que passa por P e Q; em seguida traçamos a reta L_2 que passa por $\mathcal{O}$ e $P * Q$ e encontramos o terceiro ponto que é $P + Q$. Veja a figura abaixo

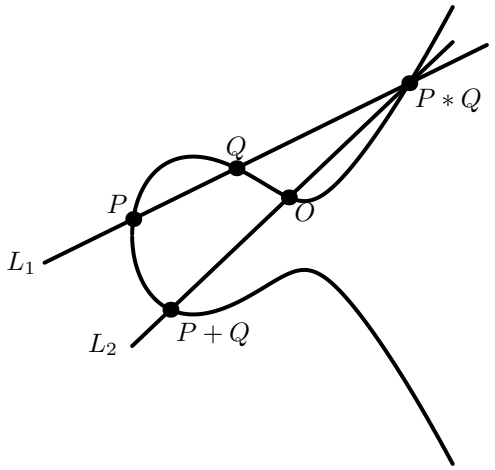


Figura 3.17

a) Da definição da lei de composição da secante-tangente a comutatividade é evidente, isto é, $P + Q = \mathcal{O} * (P * Q) = \mathcal{O} * (Q * P) = Q + P$.

b) $\mathcal{O}$ é o elemento neutro da operação $+$. De fato, se P é um ponto qualquer de C, então o ponto $P * \mathcal{O}$ é obtido passando uma reta L por P e $\mathcal{O}$. Assim, P, $\mathcal{O}$ e $P * \mathcal{O}$ estão sobre L. Para obter $\mathcal{O} * (P * \mathcal{O})$, passamos uma reta por $\mathcal{O}$ e $P * \mathcal{O}$. Como tal reta é a reta L, o terceiro ponto será P. Assim, temos $\mathcal{O} * (P * \mathcal{O}) = P$, i.e., $P + \mathcal{O} = P$. Veja a figura 3.18.

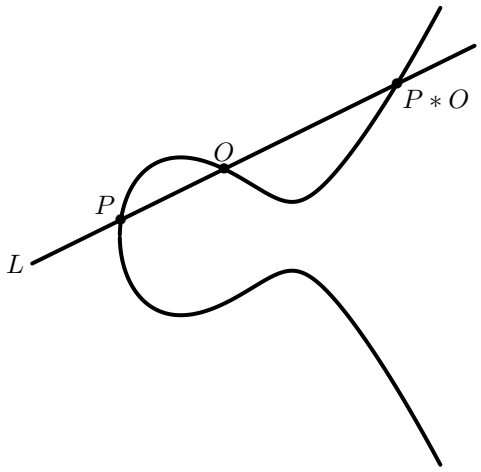


Figura 3.18

c) Como a cúbica é não singular, a reta tangente é bem definida em cada ponto. Seja, então, L a reta tangente à curva em $\mathcal{O}$. Seja S o ponto de interseção de L com a curva. Para cada ponto P na curva, ligamos P a S obtendo o ponto $P * S$. Em seguida ligamos P a $P * S$ obtendo S. Então, ligando $\mathcal{O}$ a S obtemos $\mathcal{O}$, pois a reta que passa por S e $\mathcal{O}$ é tangente à curva em $\mathcal{O}$, portanto passando uma vez em S e duas vezes em $\mathcal{O}$. Temos, então, $P + (P * S) = \mathcal{O} * [(P * (P * S)] = \mathcal{O} * S = \mathcal{O}$, donde segue que $-P = P * S$.

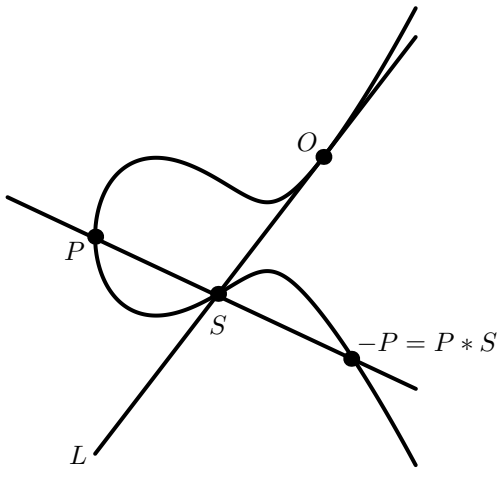


Figura 3.19

d) A operação $+$ é associativa, i.e., $(P+Q)+R = P+(Q+R)$, quaisquer que sejam P, Q e R sobre a curva. Para verificar esta igualdade, basta verificar

que $(P+Q)*R = P*(Q+R)$, pois, feito isso, aplicamos $\mathcal{O}$ a ambos os lados da igualdade e obtemos $(P+Q)+R = P+(Q+R)$.
i) Para obter $P+Q$, temos que tomar o ponto $P*Q$ e ligá-lo ao ponto $\mathcal{O}$, obtendo assim o terceiro ponto da interseção da reta com a cúbica. Depois, traçamos a reta que passa por $P+Q$ e R e encontramos o ponto $(P+Q)*R$.

ii) Para obter o ponto $P*(Q+R)$ primeiro encontramos $Q*R$ e depois o ligamos ao ponto $\mathcal{O}$, obtendo $Q+R$. Em seguida, unimos $Q+R$ a P obtendo o ponto $P*(Q+R)$. Veja a figura 3.20.

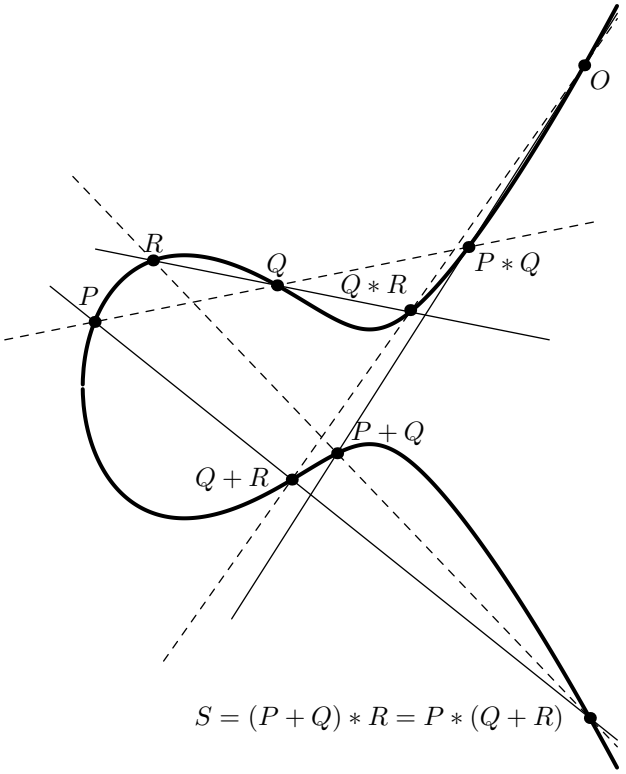


Figura 3.20

Na figura acima, seja C_1 o conjunto dos pontos situados sobre as 3 linhas pontilhadas e C_2 o conjunto dos pontos sobre as 3 linhas contínuas, que são duas cúbicas degeneradas. Assim, temos nove pontos $\mathcal{O}, P, Q, R, P*Q$, $Q*R, P+Q, P+R$ e o ponto de interseção, digamos S, das retas L_1, que passa por P e $Q+R$, e L_2, que passa por $P+Q$ e R. As duas cúbicas C_1 e C_2 passam pelos nove pontos, e a cúbica original passa por 8 desses nove pontos. Então, pelo resultado anterior temos que o nono ponto também pertence à curva C, donde concluímos que $(P+Q)*R = P*(Q+R) = S$.

3.4 Fórmulas Explícitas

3.4.1 Forma Normal de Weierstrass

Na última seção apresentamos a definição de curva elítica. Fixando um ponto $\mathcal{O}$ pertencente a uma curva elítica, definimos uma estrutura de grupo aditivo no conjunto dos pontos da curva, tendo $\mathcal{O}$ como elemento neutro. Estamos, agora, interessados em obter fórmulas explícitas para a lei (de adição) de grupo. Para fazer com que estas fórmulas tenham uma expressão mais simples, vamos considerar a equação da curva elítica numa forma especial chamada forma normal de Weierstrass.

Definição 3.4.1.1 Dizemos que a equação de uma curva elítica $C(K)$ está na forma normal de Weierstrass, se ela tem a forma

$$F(X,Y,Z) = Y^2Z + a_1XYZ + a_3YZ^2 - X^3 - a_2X^2Z - a_4XZ^2 - a_6Z^3 = 0,$$

onde $a_1, a_2, ..., a_6 \in K$, ou em coordenadas não homogêneas (fazendo $x = \frac{X}{Z}$ e $y = \frac{Y}{Z}$)

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6. \tag{3.2}$$

A equação 3.2 pode ser simplificada por meio de mudanças de coordenadas. Se a característica p do corpo K for igual a 2, a equação se reduz a uma das formas:

$$y^2 + xy = x^3 + a_2x^2 + a_6 \quad \text{ou}$$
$$y^2 + a_3y = x^3 + a_4x + a_6.$$

Para corpos de característica 3, a equação 3.2 assume uma das formas

$$y^2 = x^3 + a_2x^2 + a_6 \quad \text{ou}$$
$$y^2 = x^3 + a_4x + a_6.$$

Se $p \neq 2$, completando o quadrado do primeiro membro da equação 3.2, acima, obtemos:

$$[y - \frac{1}{2}(a_1x + a_3)]^2 = \text{cúbica em } x.$$

Substituindo $[y - \frac{1}{2}(a_1x + a_3)]^2$ por y obtemos

$$y^2 = \text{cúbica em } x.$$

Se tivermos também $p \neq 3$, então, podemos eliminar o termo quadrático do segundo membro trocando-se $x - \alpha$ por x, para um α conveniente.

Pode também acontecer que a cúbica do segundo membro tenha o coeficiente de x^3 diferente de 1. Basta então trocar x por λx e y por $\lambda^2 y$, onde λ é o coeficiente líder da cúbica do segundo membro. Assim, para valores de p diferentes de 2 e de 3, podemos supor que a equação normal de Weierstrass assume a forma simplificada $y^2 = x^3 + ax + b$.

Seja $F(x, y) = y^2 - f(x)$, e consideremos as suas derivadas parciais

$$\frac{\partial F}{\partial x} = -f'(x) \quad \text{e} \quad \frac{\partial F}{\partial y} = 2y.$$

Se $P_0 = (x_0, y_0)$ for um ponto singular sobre a curva, então

$$\frac{\partial F}{\partial x}(P_0) = \frac{\partial F}{\partial y}(P_0) = 0,$$

isto é, $y_0 = f(x_0) = 0$ e $f'(x_0) = 0$. Assim, $f(x)$ e $f'(x)$ tem uma raiz comum x_0, donde segue que x_0 é uma raiz dupla (ou tripla) de f. Reciprocamente, se f tiver uma raiz dupla (ou tripla) em x_0, então $(x_0, 0)$ será um ponto singular sobre a curva. Portanto, uma cúbica do tipo $y^2 = f(x) = x^3 + ax + b$ é uma curva elítica se, e somente se, f não tem raízes repetidas.

Sabe-se da teoria do discriminante que f não tem raízes repetidas, se e somente se, o discriminante $\Delta = -16(4a^3 + 27b^2)$ é diferente de zero [GL02]. Portanto, uma cúbica dada por $y^2 = f(x) = x^3 + ax + b$ é uma curva elítica se, e somente se, $\Delta \neq 0$. Quando $K = \mathbb{C}$ e os coeficientes de $f(x) = x^3 + ax + b$ são não nulos, temos os seguintes casos a considerar:

1) $\Delta < 0$. A equação $f(x) = 0$ tem somente uma raiz real, e o gráfico da curva tem somente uma componente conexa.

2) $\Delta > 0$. A equação $f(x) = 0$ tem três raízes reais, e o gráfico da curva tem duas componentes conexas.

3) $\Delta = 0$. A curva não é uma curva elítica, visto que tem um ponto singular.

Mostramos a seguir os possíveis gráficos para cúbicas com equações do tipo $y^2 = f(x) = x^3 + ax + b$.

$y^2 = x^3 - \frac{1}{2}x + \frac{1}{2}$
$\Delta = -100$

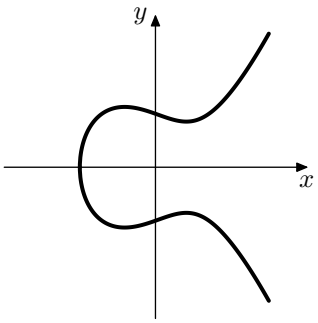


Figura 3.21

$y^2 = x^3 + x$
$\Delta = -64$

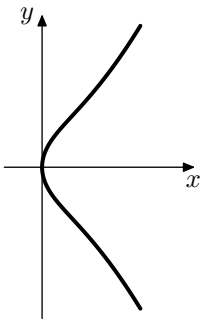


Figura 3.22

$y^2 = x^3 - x$
$\Delta = 64$

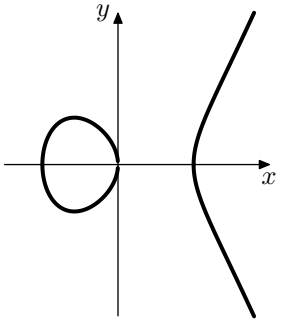


Figura 3.23

$y^2 = x^3 - 3x + 2$
$\Delta = 0$

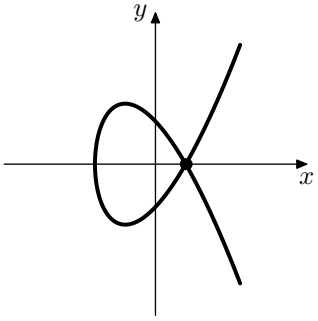


Figura 3.24

$y^2 = x^3 - 3x - 2$
$\Delta = 0$

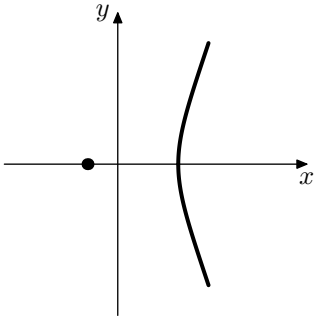


Figura 3.25

$y^2 = x^3$
$\Delta = 0$

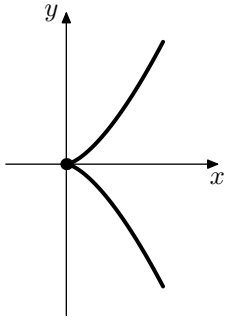


Figura 3.26

Fixado um ponto $\mathcal{O}$ em uma curva elítica $C(K)$, sempre é possível escrever a equação de $C(K)$ na forma normal de Weierstrass. Em outras palavras, pode-se provar que dada uma curva elítica $C(K)$ definida sobre um corpo K e $\mathcal{O}$ um ponto pertencente a $C(K)$, existe um isomorfismo ϕ de $C(K)$ sobre uma curva elítica $C'(K)$ dada por uma equação de Weierstrass. Além disso, $\phi(\mathcal{O}) = (0:1:0)$ e o ponto $(0:1:0)$ é o único ponto no infinito da curva $C'(K)$ (veja [Sil86]).

Para ilustrar, consideremos a cúbica de equação $u^3 + v^3 = 1$. Esta curva contém, no plano projetivo, o ponto $\mathcal{O} = (1:-1:0)$. A mudança de variáveis $(x, y) = \varphi(u, v) = \left(\frac{12}{u+v}, \frac{36(u-v)}{u+v}\right)$ associa a cada ponto da curva $u^3 + v^3 = 1$, um ponto sobre a curva $y^2 = x^3 - 432$. Calculando u e v em função de x e y, obtemos a transformação inversa dada por $(u, v) = \varphi^{-1}(x, y) = \left(\frac{36+y}{6x}, \frac{36-y}{6x}\right)$. Temos, assim, uma bijeção que a cada ponto (u, v) da curva $u^3 + v^3 = 1$ associa um único ponto (x, y) da curva $y^2 = x^3 - 432$. Se considerarmos coordenadas homogêneas, vemos que a função

$$\begin{gathered}\phi : C(K) \longrightarrow C'(K),\\ (U:V:W) \mapsto (12W : 36(U-V) : U+V)\end{gathered}$$

onde $C(K)$ é a curva $U^3 + V^3 = W^3$ (ou $u^3 + v^3 = 1$, em coordenadas não homogêneas) e $C'(K)$ é a curva $Y^2Z = X^3 - 432Z^3$ (ou $y^2 = x^3 - 432$, em coordenadas não homogêneas), é um isomorfismo que associa a cada ponto de $C(K)$, um ponto da curva $C'(K)$, e tal que $\phi(1:-1:0) = (0:1:0)$. Além disso, se $(U:V:W)$ pertence a $C(K)$ e $(U:V:W) \neq \mathcal{O}$, i.e., se $(U:V:W)$ pertence à parte afim de $C(K)$, então temos que

$$\phi(u:v:1) = \phi\left(\frac{U}{W} : \frac{V}{W} : 1\right) =$$

$$= \left(\frac{12W}{U+V} : \frac{36(U-V)}{U+V} : 1\right) = \left(\frac{12}{u+v} : \frac{36(u-v)}{u+v} : 1\right).$$

Veja as figuras abaixo.

$u^3 + v^3 = 1$

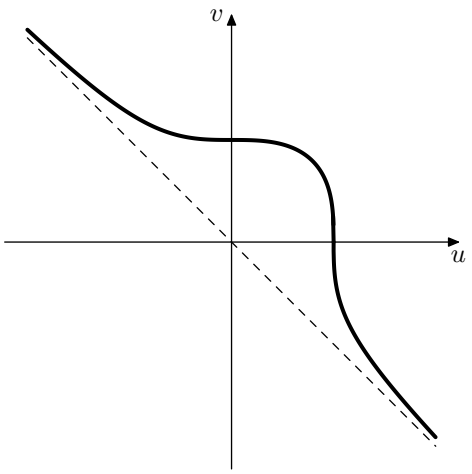


Figura 3.27

$y^2 = x^3 - 432$

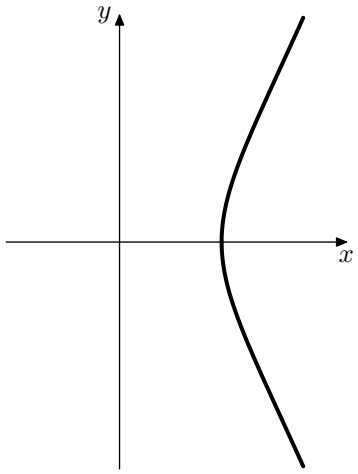


Figura 3.28

3.4.2 Fórmulas explícitas para a lei de grupo

Seja $C(K)$ uma curva elítica sobre um corpo K dada por uma equação normal de Weierstrass. Suponhamos que a característica de K é diferente de 2 e de 3. Então, a equação normal de Weierstrass é do tipo $y^2 = x^3 + ax + b$, ou em coordenadas homogêneas, $Y^2Z = X^3 + aXZ^2 + bZ^3$ $(*)$. A interseção com a reta $Z = 0$ é obtida fazendo $Z = 0$ na equação $(*)$, obtendo assim $X^3 = 0$. Vemos, então, que o ponto no infinito, $\mathcal{O} = (0 : 1 : 0)$, tem multiplicidade 3. Tal ponto no infinito é o ponto onde as retas verticais ($x =$ constante) se interceptam. Lembremos também que o discriminante $\Delta = -16(4a^3 + 27b^2)$ não é nulo.

Queremos encontrar fórmulas explícitas para a lei de grupo, isto é, dados dois pontos P_1 e P_2 pertencentes a uma cúbica dada por uma equação do tipo $y^2 = x^3 + ax + b$, queremos encontrar fórmulas que nos dê $P_1 + P_2$ em função das coordenadas de P_1 e P_2. Lembremos que para somar dois pontos, traçamos, primeiramente, a reta que passa por P_1 e P_2, encontrando assim o terceiro ponto de interseção entre a reta e a cúbica. Em seguida, traçamos a reta que passa por $\mathcal{O}$ e por $P_1 * P_2$ que é exatamente a reta vertical que passa por $P_1 * P_2$. Como uma cúbica dada por uma equação normal de Weierstrass é simétrica em relação ao eixo x, o ponto $P_1 + P_2$ será o ponto simétrico de $P_1 * P_2$ em relação a tal eixo. Os pontos $P_1 * P_2$ e $P_1 + P_2$ terão, portanto, a mesma abcissa x e as ordenadas com sinais contrários.

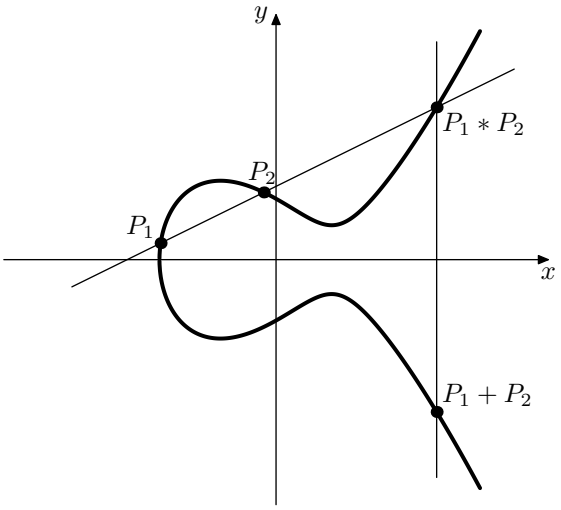


Figura 3.29

Antes de apresentar as fórmulas explícitas para a adição, observemos que dado um ponto $P = (x, y)$, o seu simétrico é o ponto $-P = (x, -y)$. De fato, a reta que passa por P e $-P$ é a reta vertical de abcissa x. Então, o terceiro ponto de interseção é o ponto $\mathcal{O}$. Assim, temos

$$P + (-P) = \mathcal{O} * [P * (-P)] = \mathcal{O} * \mathcal{O} = \mathcal{O}.$$

Se $P = \mathcal{O}$, então $-P = \mathcal{O}$, pois $\mathcal{O} + \mathcal{O} = \mathcal{O} * (\mathcal{O} * \mathcal{O}) = \mathcal{O} * \mathcal{O} = \mathcal{O}$.

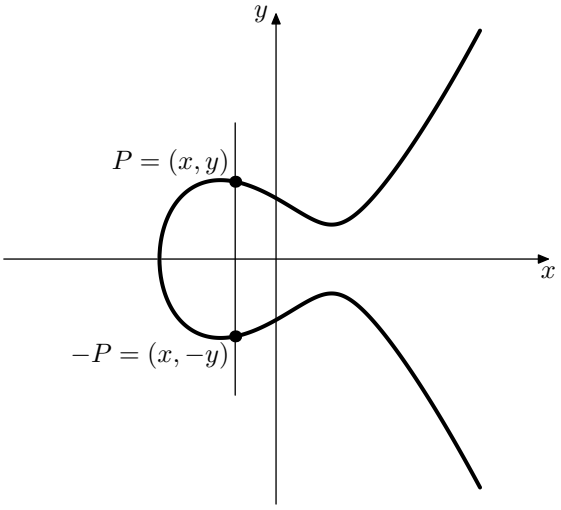


Figura 3.30

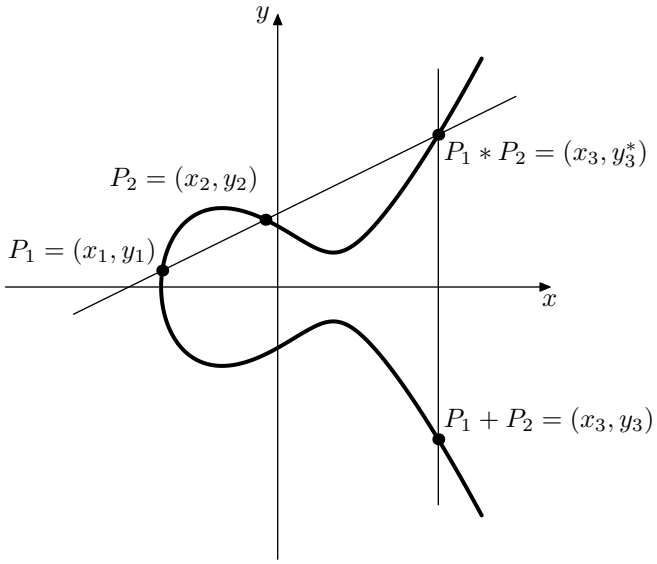


Figura 3.31

Apresentaremos, agora, fórmulas para calcular a soma P_1+P_2 de maneira explícita. Para obter estas fórmulas, vamos supor que a reta que passa pelos pontos P_1 e P_2 sobre a curva elítica $C(K)$ é uma reta não-vertical, isto é, $P_1 + P_2 \neq \mathcal{O}$.

1) Adição dos pontos P_1 e P_2, com $P_1 \neq P_2$. Façamos $P_1 = (x_1, y_1)$, $P_2 = (x_2, y_2)$, $P_1 * P_2 = (x_3, y_3^*)$ e $P_1 + P_2 = (x_3, y_3)$, onde $y_3 = -y_3^*$. A equação da reta que passa por P_1 e P_2 é dada pela equação $y = \lambda x + \nu$, onde $\lambda = \frac{y_2 - y_1}{x_2 - x_1}$, $\nu = y_1 - \lambda x_1 = y_2 - \lambda x_2$. Para obter $P_1 + P_2$, o terceiro ponto de interseção desta reta com a cúbica, substituímos o valor de $y = \lambda x + \nu$, obtendo $(\lambda x + \nu)^2 = x^3 + ax + b$. Desenvolvendo o binômio e agrupando os termos semelhantes, obtemos uma equação do 3^o grau na variável x, a saber,

$$x^3 - \lambda^2 x^2 + (a - 2\lambda\nu)x + (b - \nu^2) = 0.$$

As três raízes desta equação são x_1, x_2, x_3. Assim, temos

$$x^3 - \lambda^2 x^2 + (a - 2\lambda\nu)x + (b - \nu^2) = (x - x_1)(x - x_2)(x - x_3).$$

Comparando os coeficientes do termo em x^2 obtemos $-\lambda^2 = -(x_1 + x_2 + x_3)$, que equivale a $x_3 = \lambda^2 - x_1 - x_2$, donde segue que

$$y_3^* = \lambda x_3 + \nu = \lambda x_3 + (y_1 - \lambda x_1) = \lambda(x_3 - x_1) + y_1,$$

que é equivalente à $y_3 = \lambda(x_1 - x_3) - y_1$. Portanto, dados dois pontos $P_1 = (x_1, y_1)$ e $P_2 = (x_2, y_2)$ pertencentes à curva $C(K)$ de equação $y^2 = x^3 + ax + b$ obtemos a soma $P_1 + P_2 = (x_3, y_3)$ através das fórmulas

$$\begin{cases} \lambda = \frac{y_2 - y_1}{x_2 - x_1} \\ \\ x_3 = \lambda^2 - x_1 - x_2 \\ \\ y_3 = \lambda(x_1 - x_3) - y_1 \end{cases}$$

Vamos agora deduzir as fórmulas para $P_1 + P_2$ no caso em que $P_1 = P_2$.

2) Duplicação do ponto P_1. Façamos $P_1 = (x_1, y_1)$, onde $y_1 \neq 0$, e $2P_1 = P_1 + P_1 = (x_3, y_3)$. Seja $y = \lambda x + \nu$ a equação da reta tangente à curva em P_1. Então, para encontrar o valor de λ, derivamos implicitamente a equação $y^2 = x^3 + ax + b$, obtendo

$$\lambda = \frac{dy}{dx} = \frac{3x_1^2 + a}{2y_1}.$$

Portanto, para obter o ponto $P_3 = (x_3, y_3)$, resultado da adição de P_1 e P_1, onde $P_1 = (x_1, y_1)$, usamos as fórmulas:

$$\begin{cases} \lambda = \frac{3x_1^2 + a}{2y_1} \\ \\ x_3 = \lambda^2 - 2x_1 \\ \\ y_3 = \lambda(x_1 - x_3) - y_1 \end{cases}$$

Estas são as fórmulas básicas para a adição de pontos de uma curva elítica quando ela está na forma normal de Weierstrass.

No caso em que os pontos P_1 e P_2 estão sobre uma mesma reta vertical a soma $P_1 + P_2$ é igual ao ponto no infinito $\mathcal{O}$.

Para ilustrar, vamos apresentar um exemplo.

Exemplo 3.4.2.1 Consideremos a curva elítica de equação $y^2 = x^3 + 2x + 1$. Sejam $P_1 = (1, -2)$ e $P_2 = (0, 1)$ dois pontos sobre a curva.
i) Para calcular $P_1 + P_2$, primeiro calculamos $\lambda = \frac{y_2 - y_1}{x_2 - x_1} = \frac{1-(-2)}{0-1} = -3$.
Depois calculamos:

$$\begin{cases} x_3 = \lambda^2 - x_1 - x_2 = (-3)^2 - 1 - 0 = 8 \quad \text{e} \\ \\ y_3 = \lambda(x_1 - x_3) - y_1 = -3(1-8) - (-2) = 23. \end{cases}$$

Então, $P_1 + P_2 = (8, 23)$.

ii) Para calcular $2P_1 = P_1 + P_1$, primeiro calculamos $\lambda = \frac{3x_1^2+a}{2y_1} = \frac{3.1^2+2}{2.(-2)} = -\frac{5}{4}$. Em seguida, calculamos:

$$\begin{cases} x_3 = \lambda^2 - 2x_1 = (-\frac{5}{4})^2 - 2.1 = -\frac{7}{16} \quad \text{e} \\ \\ y_3 = \lambda(x_1 - x_3) - y_1 = (-\frac{5}{4}).(1 + \frac{7}{16}) - (-2) = \frac{13}{64}. \end{cases}$$

Assim, $2P_1 = P_1 + P_1 = (\frac{7}{16}, \frac{13}{64})$.

Consideraremos, agora, o caso de uma curva elítica sobre um corpo finito.

Exemplo 3.4.2.2 Consideremos a curva elítica $C : y^2 = x^3 + x + 4$ sobre o corpo $\mathbb{Z}_{11}$. Observe que temos nesse caso $a = 1$, $b = 4$ e $\Delta \neq 0$. Por simples verificação, vemos que os pontos $P_1 = (0, 2)$ e $P_2 = (3, 1)$ pertencem à curva C. Usando as fórmulas acima, de adição de pontos, temos

$$\begin{cases} \lambda = \frac{1-2}{3-0} = \frac{10}{3} = 7 \\ \\ x_3 = 7^2 - 0 - 3 = 5 - 0 - 3 = 2 \\ \\ y_3 = 7(0-2) - 2 = 8 - 2 = 6. \end{cases}$$

Então, $P_1 + P_2 = (0, 2) + (3, 1) = (2, 6)$ em $\mathbb{Z}_{11}$.

Vamos calcular as coordenadas do ponto $2P_2$. Usando as fórmulas de duplicação de pontos obtemos

$$\begin{cases} \lambda = \frac{3\cdot 3^2+1}{2\cdot 1} = \frac{6}{2} = 3 \\ \\ x_3 = 3^2 - 3 - 3 = 9 - 3 - 3 = 3 \\ \\ y_3 = 3(3-3) - 1 = 0 - 1 = 10. \end{cases}$$

Então, $2P_2 = 2(3,1) = (3,10)$ em $\mathbb{Z}_{11}$.
Utilizando as fórmulas acima, podemos obter todos os pontos da curva. Então, temos $C(\mathbb{Z}_{11}) = \{(0,2),(9,4),(3,1),(2,6),(2,5),(3,10),(9,7),(0,9),\mathcal{O}\}$.

O teorema abaixo dá uma estimativa para o número de pontos de uma curva elítica sobre $\mathbb{Z}_p$ (veja [Sil86]).

Teorema 3.4.2.1 (Teorema de Hasse) Seja p um número primo e $C(\mathbb{Z}_p)$ uma curva elítica sobre $\mathbb{Z}_p$. Então, $p+1-2\sqrt{p} \leq \#C(\mathbb{Z}_p) \leq p+1+2\sqrt{p}$.

Para a curva elítica do exemplo acima, segue do teorema de Hasse que

$$5,37 \approx 11+1-2\sqrt{11} \leq \#C(\mathbb{Z}_{11}) = 9 \leq 11+1+2\sqrt{11} \approx 18,63.$$

3.5 Redução Módulo p

Nesta seção estudaremos uma função, chamada de redução módulo p, que leva pontos de $\mathbb{P}^2(\mathbb{Q})$ em pontos de $\mathbb{P}^2(\mathbb{Z}_p)$. Observemos primeiramente que a função

$$\begin{array}{ccc} \mathbb{Z} & \longrightarrow & \mathbb{Z}_p \ , \\ z & \longmapsto & \tilde{z} \end{array}$$

é um homomorfismo de anéis ($\tilde{z}$ é a classe de equivalência de z em $\mathbb{Z}_p$ que identificamos com $z \mod p$, o seu resíduo módulo p).

Dizemos que uma terna homogênea de coordenadas $(A:B:C)$ em $\mathbb{P}^2(\mathbb{Q})$ está normalizada, se A, B e C são inteiros sem fatores primos comuns. Assim, cada ponto $P \in \mathbb{P}^2(\mathbb{Q})$ tem uma terna normalizada de coordenadas que é única a menos de sinal. Para obtê-la começamos com uma terna de coordenadas racionais, multiplicamo-la pelo *mmc* dos denominadores e depois dividimos a terna resultante pelo *mdc* de suas coordenadas. Por exemplo, a partir da terna $\left(\frac{5}{12}:\frac{15}{8}:\frac{25}{4}\right)$, obtemos a terna $(10:45:150)$ multiplicando a primeira pelo $mmc(12,8,4)$, que é igual a 24. Em seguida, dividimos esta nova terna por 5, que é o $mdc(10,45,150)$, obtendo $(2:9:30)$.

Seja agora p um número primo fixo. Para cada $m \in \mathbb{Z}$ seja $\tilde{m} \in \mathbb{Z}_p$ o seu resíduo módulo p. Se $(l:m:n)$ é uma terna normalizada de coordenadas de um ponto $P \in \mathbb{P}^2(\mathbb{Q})$, então a terna $(\tilde{l}:\tilde{m}:\tilde{n})$ define um ponto em $\mathbb{P}^2(\mathbb{Z}_p)$ pois pelo menos um dos três números l, m e n não é divisível por p, e portanto $(\tilde{l},\tilde{m},\tilde{n}) \neq (0,0,0)$, i.e., pelo menos uma das coordenadas não é zero em $\mathbb{Z}_p$. Como P determina a terna $(l:m:n)$ a menos de sinal, o ponto $\widetilde{P}$ depende só de P e não da escolha das coordenadas de P. Assim, a função $P \to \widetilde{P}$

de $\mathbb{P}^2(\mathbb{Q})$ em $\mathbb{P}^2(\mathbb{Z}_p)$ está bem definida e é chamada, por razões óbvias, de redução módulo p.

Por exemplo, o ponto $P = (2 : 9 : 30)$ em $\mathbb{P}^2(\mathbb{Q})$ é levado no ponto $\widetilde{P} = (2 : 4 : 0)$ em $\mathbb{P}^2(\mathbb{Z}_5)$. Se considerarmos a outra terna normalizada $P = (-2 : -9 : -30)$, obteremos $\widetilde{P} = (3 : 1 : 0)$. Mas ambas as ternas obtidas representam o mesmo ponto de $\mathbb{P}^2(\mathbb{Z}_5)$, pois $\widetilde{P} = (2 : 4 : 0) = 4(3 : 1 : 0)$.

Lembremos que estamos considerando que os pontos no infinito em $\mathbb{P}^2(\mathbb{Q})$ são os pontos com a terceira coordenada igual a zero, isto é, os pontos da forma $(a : b : 0)$, com $a, b \in \mathbb{Q}$. O ponto $(a : b : 0) \in \mathbb{P}^2(\mathbb{Q})$ é levado em $(\tilde{a} : \tilde{b} : 0)$ que é um ponto no infinito em $\mathbb{P}^2(\mathbb{Z}_p)$. Serão os pontos da forma $(a : b : 0)$ os únicos pontos em $\mathbb{P}^2(\mathbb{Q})$ que são levados em pontos no infinito de $\mathbb{P}^2(\mathbb{Z}_p)$? A resposta é não. Por exemplo, se tomarmos o ponto $P = (\frac{1}{p} : 0 : 1) = (1 : 0 : p)$ vemos que $\widetilde{P} = (1 : 0 : 0)$ é um ponto no infinito em $\mathbb{P}^2(\mathbb{Z}_p)$. Portanto, a função redução módulo p, não leva $\mathbb{A}^2(\mathbb{Q})$ em $\mathbb{A}^2(\mathbb{Z}_p)$. Resumindo, consideremos um ponto $P = (a : b : c)$ dado em coordenadas normalizadas. Dois casos podem ocorrer:

i) Se $p|c$, c é múltiplo de p e portanto $(\tilde{a} : \tilde{b} : \tilde{c}) = (\tilde{a} : \tilde{b} : 0)$ é um ponto no infinito em $\mathbb{P}^2(\mathbb{Z}_p)$.

ii) Se $p \nmid c$, então $(\tilde{a} : \tilde{b} : \tilde{c})$ não é um ponto no infinito em $\mathbb{P}^2(\mathbb{Z}_p)$, isto é, $(\tilde{a} : \tilde{b} : \tilde{c}) \in \mathbb{A}^2(\mathbb{Z}_p)$ (pois $\tilde{c} \neq 0$ em $\mathbb{Z}_p$).

De i) e ii) vemos que dado $P = (a : b : c) \in \mathbb{P}^2(\mathbb{Q})$, $\widetilde{P} = (\tilde{a} : \tilde{b} : \tilde{c})$ é um ponto no infinito em $\mathbb{P}^2(\mathbb{Z}_p)$ se, e somente se, $p|c$.
Vemos então que os pontos de $\mathbb{P}^2(\mathbb{Q})$, dados em coordenadas normalizadas, cuja redução módulo p são os pontos no infinito de $\mathbb{P}^2(\mathbb{Z}_p)$, são os pontos no infinito de $\mathbb{P}^2(\mathbb{Q})$ (os pontos da forma $(a : b : 0)$), ou os pontos de $\mathbb{A}^2(\mathbb{Q})$ que têm a terceira coordenada divisível por p (os pontos $(a : b : c)$ com $p|c$).

Redução módulo p

$$\mathbb{P}^2(\mathbb{Q}) \longrightarrow \mathbb{P}^2(\mathbb{Z}_p)$$

$$\{(a : b : c) \text{ normalizada, com } p \nmid c\} \subset \mathbb{A}^2(\mathbb{Q}) \quad \longrightarrow \quad \{(\tilde{a} : \tilde{b} : \tilde{c})\} \subset \mathbb{A}^2(\mathbb{Z}_p)$$

$$\left\{ \begin{array}{c} \{(a : b : c) \text{ norm., com } p|c, c \neq 0\} \subset \mathbb{A}^2(\mathbb{Q}) \\ \\ \{ \text{ pontos no infinito } \} \subset \mathbb{P}^2(\mathbb{Q}) \end{array} \right\} \longrightarrow \text{pontos inf. em } \mathbb{P}^2(\mathbb{Z}_p)$$

Seja, agora, $P = \left(\frac{a_1}{c_1}, \frac{b_1}{c_2}\right) = \left(\frac{a_1}{c_1} : \frac{b_1}{c_2} : 1\right)$ um ponto de $\mathbb{A}^2(\mathbb{Q})$ onde $\frac{a_1}{c_1}$ e $\frac{b_1}{c_2}$ são frações irredutíveis, i.e., o mdc entre o numerador e o denominador

de cada fração é igual a 1. Como vimos anteriormente, podemos escrever P na forma normalizada, digamos $P = (a : b : c)$ com a, b e c sem fator primo comum. Temos que $p \mid c$ se , e somente se, pelo menos um dos números c_1 e c_2 têm o fator primo p na sua decomposição. Portanto, a redução módulo p de um ponto $P \in \mathbb{A}^2(\mathbb{Q})$ será um ponto no infinito em $\mathbb{P}^2(\mathbb{Z}_p)$ se, e somente se, pelo menos um dos denominadores das coordenadas afins de P, c_1 ou c_2, tiver um fator p.

Consideremos agora uma curva $C(\mathbb{Q}) : F(X, Y, Z) = 0$ em $\mathbb{P}^2(\mathbb{Q})$, i.e., uma curva tal que os coeficientes do polinômio F, associado a ela, são números racionais. Podemos supor que os coeficientes de F são números inteiros com máximo divisor comum igual a 1; para isto basta eliminar os denominadores dos coeficientes racionais de F e depois dividir os coeficientes obtidos pelo seu *mdc*. Para este novo polinômio F, normalizado, consideremos $\widetilde{F}$, o polinômio obtido pela redução módulo p dos coeficientes de F. Este polinômio $\widetilde{F}$, assim obtido, é não-nulo e define uma curva $\widetilde{C}(\mathbb{Z}_p)$ em $\mathbb{P}^2(\mathbb{Z}_p)$. Agora, se $(l : m : n)$ é uma terna normalizada de coordenadas e se $F(l, m, n) = 0$, então $\widetilde{F}(\tilde{l}, \tilde{m}, \tilde{n}) = 0$ pois a função $x \to \tilde{x}$ de $\mathbb{Z}$ em $\mathbb{Z}_p$ é um homomorfismo de anéis. Em outras palavras, se P é um ponto sobre $C(\mathbb{Q})$, então $\widetilde{P}$ é um ponto sobre $\widetilde{C(\mathbb{Z}_p)}$. Logo, a redução módulo p leva $C(\mathbb{Q})$ em $\widetilde{C}(\mathbb{Z}_p)$. Por exemplo, seja $C(\mathbb{Q})$ a curva dada por $13X^2Y+10XYZ+21Y^2Z = 0$. Então, fazendo a redução módulo 5, obtemos a curva $\widetilde{C} : 3X^2Y + Y^2Z = 0$ em $\mathbb{P}^2(\mathbb{Z}_5)$.

Vamos mostrar que a redução módulo p de $C(\mathbb{Q})$ em $\widetilde{C}(\mathbb{Z}_p)$, com as operações de grupo (soma de pontos) de $C(\mathbb{Q})$ e de $\widetilde{C}(\mathbb{Z}_p)$ é um homomorfismo de grupos. Sabemos que dados uma função f e dois conjuntos quaisquer A e B, contidos no domínio de f, vale a seguinte inclusão $f(A \cap B) \subset f(A) \cap f(B)$, isto é, a imagem direta da interseção de dois conjuntos está contida na interseção das suas imagens diretas. Aplicando este resultado à função de redução módulo p e aos conjuntos dos pontos racionais de duas curvas $C_1(\mathbb{Q})$ e $C_2(\mathbb{Q})$ temos

$$(\widetilde{C_1(\mathbb{Q}) \cap C_2(\mathbb{Q})}) \subset (\widetilde{C_1}(\mathbb{Z}_p) \cap \widetilde{C_2}(\mathbb{Z}_p)).$$

Observemos que os graus das curvas reduzidas $\widetilde{C_i}$ são os mesmos das curvas C_i, pois os polinômios associados a elas são homogêneos. Então, pelo Teorema de Bezout a interseção antes e depois da redução módulo p tem o mesmo número de pontos, desde que contemos multiplicidades e, além disso, que as reduzidas não tenham componentes em comum. Mas o Teorema de Bezout exige que o corpo em que se está trabalhando seja algebricamente fechado, que não é o caso do corpo dos números racionais. No entanto, se

supusermos que todos os pontos de interseção complexos são racionais, isto é, todos os pontos de interseção das curvas são racionais mesmo que olhemos os pontos com coordenadas em $\mathbb{C}$, o teorema poderá ser aplicado.

No que se segue, consideraremos o caso particular em que uma das curvas é uma cúbica e a outra é uma reta. Estudar este caso é suficiente para o que nos interessa que é a aplicação desse resultado à adição definida no conjunto dos pontos de uma curva elítica. Além disso, este caso é fácil de demonstrar e será apresentado na proposição 3.5.0.1. Antes de demonstrar a proposição, vamos apresentar alguns resultados que serão utilizados em sua demonstração.

Lema 3.5.0.1 Seja (a, b, c) uma terna de números inteiros com $mdc(a, b, c) = 1$. Então, existe uma matriz 3×3 com coeficientes inteiros, determinante igual a 1 e com $3^{\underline{a}}$ linha igual a (a, b, c).

Prova: Seja $d = mdc(b, c)$. Sejam r e s inteiros tais que $rc - sb = d$. Visto que $mdc(a, d) = 1$, podemos escolher inteiros t e u tais que $td + ua = 1$. Da escolha de r e s temos que $mdc(r, s) = 1$ $\left(r\left(\frac{c}{d}\right) - s\left(\frac{b}{d}\right) = 1 \Leftrightarrow mdc(r, s) = 1\right)$ e portanto podemos escolher inteiros v e w tais que $vs - wr = u$, pois $\frac{c}{d}$ e $\frac{b}{d}$ são inteiros. Então, a matriz

$$\begin{pmatrix} t & v & w \\ 0 & r & s \\ a & b & c \end{pmatrix}$$

tem, claramente, as propriedades desejadas.

□

Lema 3.5.0.2 Seja L uma reta em $\mathbb{P}^2(\mathbb{Q})$ dada pela equação

$$L : aX + bY + cZ = 0,$$

onde $(a : b : c)$ é uma terna normalizada. Então, existe uma transformação linear

$$T : \mathbb{P}^2(\mathbb{Q}) \longrightarrow \mathbb{P}^2(\mathbb{Q}),$$

compatível com a redução módulo p que leva L na reta no infinito $L' : Z' = 0$.

Prova: Pelo lema anterior, existe uma matriz formada por números inteiros, determinante igual a 1 e terceira linha (a, b, c). Denotemos por (t_{ij}) tal matriz. A matriz $(m_{ij}) = (t_{ij})^{-1}$, matriz inversa da matriz (t_{ij}), também será formada por elementos inteiros. As matrizes reduzidas módulo p $(\widetilde{t_{ij}})$ e

$(\tilde{m}_{ij})$ são, portanto, uma inversa da outra. Seja, então, T a transformação linear dada por

$$T\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix} \cdot \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix}.$$

A transformação T, assim definida, tem as propriedades requeridas.

□

Proposição 3.5.0.1 Sejam C uma cúbica irredutível e não-singular e L uma reta definidas em $\mathbb{P}^2(\mathbb{Q})$. Se $C \cap L = \{P_1, P_2, P_3\}$ em coordenadas normalizadas, e se $\widetilde{L}$ não é uma componente de $\widetilde{C}$, então $\widetilde{C} \cap \widetilde{L} = \{\widetilde{P_1}, \widetilde{P_2}, \widetilde{P_3}\}$.

Prova: **i)** Suponhamos que L seja a reta no infinito $Z = 0$. Seja $F(X, Y, Z) = 0$ uma equação normalizada para C. Como $\widetilde{L}$ não é componente de $\widetilde{C}$, isto é, $\widetilde{F}(X, Y, Z)$ não é da forma $Z \cdot \widetilde{P}(X, Y, Z)$, o polinômio $\widetilde{F}(X, Y, 0)$ não é identicamente nulo, que equivale a dizer que pelo menos um dos coeficientes de $F(X, Y, 0)$ não é divisível por p.

Para cada ponto de interseção P_i, seja $P_i = (l_i : m_i : 0)$ dado em coordenadas normalizadas. Como $F(X, Y, 0) = 0$ é um polinômio homogêneo em X e Y temos que

$$F(X, Y, 0) = c\prod_{i=1}^{3}(m_i X - l_i Y) \tag{3.3}$$

para alguma constante c. Visto que cada um dos polinômios lineares do lado direito de (2.1) está normalizado, e algum coeficiente de F não é divisível por p, vemos que c deve ser um inteiro não divisível por p. Então podemos reduzir módulo p para obter

$$\widetilde{F}(X, Y, 0) = \tilde{c}\prod_{i=1}^{3}(\tilde{m}_i X - \tilde{l}_i Y), \tag{3.4}$$

o que mostra que $\widetilde{C} \cap \widetilde{L} = \{\widetilde{P}_1, \widetilde{P}_2, \widetilde{P}_3\}$ como queríamos.

ii) Se L não for a reta no infinito, então, usamos a transformação definida pelo lema 3.5.0.2 e caímos no caso anterior.

□

Proposição 3.5.0.2 Sejam $C(\mathbb{Q})$ uma cúbica irredutível e não-singular em $\mathbb{P}^2(\mathbb{Q})$ e seja $\mathcal{O}$ um ponto em $C(\mathbb{Q})$ tomado como elemento neutro da estrutura de grupo sobre $C(\mathbb{Q})$. Suponha que $\widetilde{C}(\mathbb{Z}_p)$ é não-singular e tomemos $\widetilde{\mathcal{O}}$ como sendo o elemento neutro para a estrutura de grupo sobre $\widetilde{C}(\mathbb{Z}_p)$. Então, a redução módulo p, $P \longrightarrow \widetilde{P}$, de $C(\mathbb{Q})$ em $\widetilde{C}(\mathbb{Z}_p)$ é um homomorfismo de grupo, i.e., $\widetilde{P+Q} = \widetilde{P} + \widetilde{Q}$, $\forall P, Q \in C(\mathbb{Q})$.

Prova: Sejam P e Q dois pontos em $C(\mathbb{Q})$ e $R = P + Q$. Sejam L_1 a reta em $\mathbb{P}^2(\mathbb{Q})$ que passa por P e Q, e L_2 a reta em $\mathbb{P}^2(\mathbb{Q})$ que passa por $P * Q$ e R. Então, temos

$$C \cap L_1 = \{P, Q, P * Q\} \quad \text{e} \quad C \cap L_2 = \{\mathcal{O}, P * Q, R\}.$$

Da proposição 3.5.0.1 segue que

$$\widetilde{C} \cap \widetilde{L_1} = \{\widetilde{P}, \widetilde{Q}, \widetilde{P * Q}\} \quad \text{e} \quad \widetilde{C} \cap \widetilde{L}_2 = \{\widetilde{\mathcal{O}}, \widetilde{P * Q}, \widetilde{R}\}.$$

Assim, temos que

$$\widetilde{R} = \widetilde{\mathcal{O}} * (\widetilde{P * Q}) = \widetilde{\mathcal{O}} * (\widetilde{P} * \widetilde{Q}) = \widetilde{P} + \widetilde{Q}.$$

□

Capítulo 4

Método de Fatoração das Curvas Elíticas

O objetivo deste capítulo é descrever um método de fatoração de números inteiros que depende do uso de curvas elíticas, o Método das Curvas Elíticas (Elliptic Curve Method - ECM) [Len87] devido a H. W. Lenstra. Tal método é inspirado em um método de fatoração sobre $\mathbb{Z}_p^*$, o método $p-1$ de Pollard [Pol74]. No parágrafo 4.1 apresentamos alguns algoritmos que são utilizados nos métodos de Pollard e de Lenstra. No parágrafo 4.2 apresentamos o método de Pollard e no parágrafo 4.3 o método de fatoração das curvas elíticas.

4.1 Algoritmos básicos

4.1.1 Algoritmo exponenciação modular

Este algoritmo é utilizado para calcular eficientemente $a^k \bmod n$, onde a, k e n são inteiros, ou dito de outra maneira, a k-ésima potência do elemento a, do grupo $\mathbb{Z}_n$ com a operação de multiplicação módulo n. Apresentaremos o algoritmo no caso mais geral onde a pertence a um grupo $(G, \cdot)$ e k é um inteiro positivo. Podemos calcular a k-ésima potência de um elemento a, multiplicando $a \cdot a$, obtendo a^2, depois encontramos $a^3 = a^2 \cdot a$ e assim sucessivamente, até obter a^k. Este algoritmo, chamado de algoritmo trivial de multiplicação, requer $k-1$ multiplicações. Porém para valores de k muito grandes, ele não é eficiente. É possível, no entanto, calcular a^k de maneira bem mais eficiente. Para calcular a^{16}, por exemplo, utilizamos 15 multiplicações usando o algoritmo trivial. Podemos, entretanto, calcular a^{16} com apenas 4 multiplicações. Calculamos $a^2 = a \cdot a$, $a^4 = a^2 \cdot a^2$, $a^8 =$

$a^4 \cdot a^4$, e finalmente $a^{16} = a^8 \cdot a^8$. Observemos que este exemplo é um caso bem particular, pois o expoente 16 é uma potência de 2. No entanto, podemos calcular qualquer potência inteira de um número, partindo desta idéia e fazendo uma pequena modificação. Para calcular a^{25}, por exemplo, calculamos a^2, a^4, a^8, a^{16}, como anteriormente, e em seguida, obtemos a^{25} fazendo $a \cdot a^8 \cdot a^{16}$. Gastamos, assim, 6 multiplicações, que é bem melhor que as 24 necessárias pelo método trivial. Olhando atentamente, o que fizemos foi calcular as várias potências de a cujos expoentes eram potências de 2, e depois multiplicar algumas delas. Mais precisamente, encontramos a representação binária de 25 e, em seguida, obtivemos a^{25} calculando o produto $a \cdot a^8 \cdot a^{16}$, onde os expoentes são as potências de 2 da representação binária de 25 que têm coeficientes iguais a 1.

$$25 = 1 + 8 + 16 = 1 \cdot 2^0 + 0 \cdot 2^1 + 0 \cdot 2^2 + 1 \cdot 2^3 + 1 \cdot 2^4$$

$$a^{25} = a \cdot a^8 \cdot a^{16} = a \cdot a^{2^3} \cdot a^{2^4}$$

Consideremos, agora, o caso geral onde k é um inteiro positivo qualquer. Para calcular a^k, o primeiro passo é obter a representação binária de k,

$$k = k_0 + k_1 \cdot 2 + k_2 \cdot 2^2 + \ldots + k_r \cdot 2^r,$$

onde r é o número inteiro, tal que $2^r \leq k < 2^{r+1}$ e k_i é 0 ou 1 ($k_r \neq 0$). A seguir, fazemos $A_0 = a$ e calculamos $A_i = a^{2^i}$, $1 \leq i \leq r$, elevando ao quadrado, repetidamente, i vezes.

$$\begin{array}{ccccc} A_1 & = & A_0 \cdot A_0 & = & a^2 \\ A_2 & = & A_1 \cdot A_1 & = & a^{2^2} \\ A_3 & = & A_2 \cdot A_2 & = & a^{2^3} \\ \vdots & & \vdots & & \vdots \\ A_r & = & A_{r-1} \cdot A_{r-1} & = & a^{2^r} \end{array}$$

Finalmente, obtemos a^k, multiplicando todos os números $A_i = a^{2^i}$ para os quais $k_i = 1$.

$$a^k = a^{k_0 + k_1 \cdot 2 + k_2 \cdot 2^2 + \ldots + k_r \cdot 2^r} = a^{k_0} \cdot (a^2)^{k_1} \cdot (a^{2^2})^{k_2} \cdots (a^{2^r})^{k_r} = \prod_{k_i = 1} a^{2^i}.$$

Para calcular a^k, utilizamos r operações de multiplicação para obter os números $A_i, 1 \leq i \leq r$, e no máximo $r - 1$ (aproximadamente r) para obter $a^k = A_1 \cdot A_2 \cdot \cdots A_r$, gastando, assim, no máximo $2r$ operações de multiplicação. Mas $2^r \leq k < 2^{r+1}$, donde segue que $r \leq \log_2 k < r + 1$. Necessitamos, assim, de no máximo $2\lfloor \log_2 k \rfloor$ multiplicações. Se $k = 1000$,

gastaremos aproximadamente 1000 multiplicações pelo algoritmo trivial e no máximo 19 usando o método acima exposto, chamado de "método dos quadrados repetidos".

Utilizando esta idéia, apresentamos o algoritmo abaixo.

Algoritmo 4.1.1.1 (Algoritmo exponenciação modular 1) Dados um elemento a de um grupo $(G, \cdot)$ e k um inteiro positivo com representação binária $k = \sum k_i 2^i$, este algoritmo calcula a^k. Em particular, se $a \in \mathbb{Z}_n$, a saída é $a^k \bmod n$.

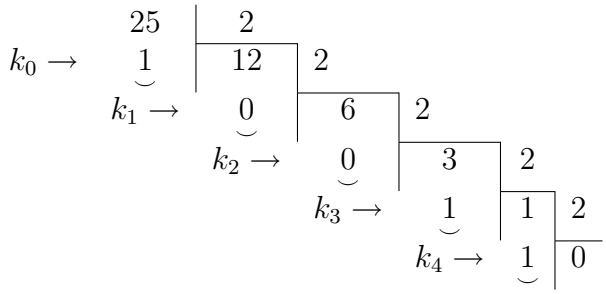
1. $b \leftarrow 1$ e $A \leftarrow a$.
2. se $k_0 = 1$, então $b \leftarrow a$
3. para i de 1 até r faça
   - 3.1 $A \leftarrow A^2$
   - 3.2 se $k_i = 1$, então $b \leftarrow bA$
4. retorne b.

Fazendo algumas modificações, obtemos o algoritmo a seguir que encontra os valores k_i e já os utiliza simultaneamente para calcular a^k. O que fazemos, é dividir k por 2 sucessivamente. Cada resto obtido é um valor para k_i. Obtemos sucessivamente $k_0, k_1, k_2, \ldots, k_r$. Simultaneamente, calculamos os valores de a^{2^i} por quadrados repetidos. Para cada $k_i = 1$ multiplicamos o produto acumulado $b = A_0 \cdot A_1 \cdot A_2 \cdot A_{i-1}$ por A_i. Para cada $k_i = 0$, apenas repetimos o valor de $b = A_0 \cdot A_1 \cdot A_2 \cdot A_{i-1}$.

Algoritmo 4.1.1.2 (Algoritmo exponenciação modular 2)
Dados um elemento a de um grupo $(G, \cdot)$ e k um inteiro positivo este algoritmo calcula a^k. Em particular, se $a \in \mathbb{Z}_n$, a saída é $a^k \bmod n$.

1. $b \leftarrow 1$ e $A \leftarrow a$.
2. enquanto $k \neq 0$, faça
   - 2.1 se k é ímpar, então $b \leftarrow b \cdot A$
   - 2.2 $k \leftarrow \lfloor \frac{k}{2} \rfloor$
   - 2.3 se $k \neq 0$, então $A \leftarrow A^2$
3. retorne b.

Para ilustrar como o algoritmo funciona tomamos como exemplo o cálculo de a^{25}. Encontramos a representação binária de 25 dividindo-o por dois, repetidamente, até encontrar quociente zero.



Procedemos como explicado acima e obtemos a tabela seguinte.

i	k	k_i	A	b
0	25	1	a	a
1	12	0	a^2	a
2	6	0	a^4	a
3	3	1	a^8	$a \cdot a^8 = a^9$
4	1	1	a^{16}	$a^9 \cdot a^{16} = a^{25}$
5	0			

4.1.2 Algoritmo euclidiano

O máximo divisor comum de dois inteiros a e b pode ser calculado a partir da decomposição deles em fatores primos. Calcula-se o conjunto de todos os divisores de a e o conjunto de todos os divisores de b; o $mdc(a, b)$ é o maior número inteiro contido na interseção destes dois conjuntos. Contudo, o cálculo do $mdc(a, b)$ feito dessa maneira não nos dá um algoritmo eficiente, pois o problema da fatoração de inteiros é relativamente difícil (é um problema matemático intratável).

O algoritmo euclidiano, ou algoritmo de Euclides, apresentado a seguir, é um algoritmo eficiente para calcular o máximo divisor comum de dois inteiros que não necessita da fatoração desses inteiros. Ele utiliza a proposição 2.2.0.6, que assegura que $mdc(a, b) = mdc(b, r)$, onde $a = bq + r$. Segue desta proposição que o problema de achar o $mdc(a, b)$ reduz-se a achar o $mdc(b, r)$, onde $a = bq + r$. Então, para calcular o máximo divisor comum entre dois inteiros positivos a e b, com $a \geq b$, dividimos a por b achando o resto r_1. Se $r_1 \neq 0$, dividimos b por r_1, obtendo o resto r_2. Se $r_2 \neq 0$, dividimos r_1 por r_2, obtendo o resto r_3. E assim por diante. Escrevemos, a seguir, a seqüência das equações obtidas por $n + 1$ divisões com resto.

$$\begin{array}{lcll} a & = & bq_1 + r_1, & 0 \leq r_1 < b \\ b & = & r_1q_2 + r_2, & 0 \leq r_2 < r_1 \\ r_1 & = & r_2q_3 + r_3, & 0 \leq r_3 < r_2 \\ r_2 & = & r_3q_4 + r_4, & 0 \leq r_4 < r_3 \\ & \vdots & & \vdots \\ r_{n-2} & = & r_{n-1}q_n + r_n, & 0 \leq r_n < r_{n-1} \\ r_{n-1} & = & r_nq_{n+1} + r_{n+1}, & 0 \leq r_{n+1} < r_n \end{array} \tag{4.1.2}$$

Como $b > r_1 > r_2 > r_3 > \ldots$, e os $r_i's$ são inteiros não negativos, algum dos restos deverá ser igual a zero, digamos r_{n+1}. Segue da proposição 2.2.0.6 que

$$mdc(a, b) = mdc(b, r_1) = \ldots = mdc(r_{n-2}, r_{n-1}) = mdc(r_{n-1}, r_n).$$

Como $r_n \mid r_{n-1}$, pois $r_{n+1} = 0$, temos que $mdc(r_{n-1}, r_n) = r_n$; logo $mdc(a, b) = r_n$. Vimos, assim, que nesse processo, o $mdc(a, b)$ é o último resto diferente de zero.

Exemplo 4.1.2.1 $mdc(2771, 1003) = 17$.

$$\begin{array}{rcrcr} 2771 & = & 2 \cdot 1003 & + & 765 \\ 1003 & = & 1 \cdot 765 & + & 238 \\ 765 & = & 3 \cdot 238 & + & 51 \\ 238 & = & 4 \cdot 51 & + & 34 \\ 51 & = & 1 \cdot 34 & + & 17 \\ 34 & = & 2 \cdot 17 & + & 0 \end{array}$$

Costumamos dispor os números que aparecem no processo no seguinte esquema

	2	1	3	4	1	2
2771	1003	765	238	51	34	17
765	238	51	34	17	0	

Algoritmo 4.1.2.1 (Algoritmo euclidiano) Determina o máximo divisor comum entre dois inteiros positivos a e b.

1. $x \leftarrow a, y \leftarrow b$
2. enquanto $r \neq 0$ faça
   - 2.1 $r \leftarrow x \mod y$
   - 2.2 $x \leftarrow y$
   - 2.3 $y \leftarrow r$
3. retorne x

4.1.3 Algoritmo euclidiano estendido

O algoritmo de Euclides pode ser estendido de forma a calcular, não somente o máximo divisor comum d de dois inteiros a e b, como também inteiros x e y tais que $ax+by=d$. A idéia é expressar cada resto r_i, obtido na seqüência de divisões 4.1.2 que fizemos para calcular o $mdc(a,b)$, em função de a e b, de maneira semelhante à fórmula $d=ax+by$. Por exemplo, para $a=bq_1+r_1$, obtemos $r_1=a(1)+b(-q_1)$. Para $b=r_1q_2+r_2$, obtemos

$$\begin{aligned} r_2 &= b-q_2r_1 = b-q_2(a-bq_2) \\ &= b-aq_2+bq_1q_2 = a(-q_2)+b(1+q_1q_2), \end{aligned}$$

e assim por diante. Calculando o $mdc(a,b)$, obtemos a seqüência de divisões (como anteriormente). Ao lado de cada equação, escrevemos as expressões dos restos r_i, onde x_i e y_i são inteiros a determinar.

$$\begin{array}{rcl@{\quad}l rcl} a & = & bq_1+r_1 & \text{e} & r_1 & = & ax_1+by_1 \\ b & = & r_1q_2+r_2 & \text{e} & r_2 & = & ax_2+by_2 \\ r_1 & = & r_2q_3+r_3 & \text{e} & r_3 & = & ax_3+by_3 \\ r_2 & = & r_3q_4+r_4 & \text{e} & r_4 & = & ax_4+by_4 \\ & \vdots & & & \vdots & & \vdots \\ r_{n-2} & = & r_{n-1}q_n+r_n & \text{e} & r_n & = & ax_n+by_n \\ r_{n-1} & = & r_nq_{n+1}+r_{n+1} & \text{e} & r_{n+1} & = & ax_{n+1}+by_{n+1} \end{array} \tag{4.1.3}$$

Podemos por a informação contida em 4.1.3 em uma tabela. Para montar a tabela, copiamos os valores referentes às n primeiras linhas de 4.1.3 e acrescentamos duas linhas, no início. Essas linhas são necessárias para iniciar o processo.

restos	quocientes	x	y
a	$*$	x_{-1}	y_{-1}
b	$*$	x_0	y_0
r_1	q_1	x_1	y_1
r_2	q_2	x_2	y_2
r_3	q_3	x_3	y_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$
r_{n-2}	q_{n-2}	x_{n-2}	y_{n-2}
r_{n-1}	q_{n-1}	x_{n-1}	y_{n-1}
r_n	q_n	x_n	y_n

(4.1.4)

As duas primeiras colunas, nós as preenchemos usando o algoritmo de Euclides. Queremos, então, descobrir como preencher as colunas 3 e 4. Para tanto, vamos supor a tabela preenchida até a $(j-1)$ - ésima linha. Escrevemos as linhas de ordem $(j-2)$, $(j-1)$ e j na tabela abaixo.

restos	quocientes	x	y
.....			
r_{j-2}	q_{j-2}	x_{j-2}	y_{j-2}
r_{j-1}	q_{j-1}	x_{j-1}	y_{j-1}
r_j	q_j	x_j	y_j
.....			

(4.1.5)

Para preencher a j - ésima linha, dividimos r_{j-2} por r_{j-1} e obtemos, pelo algoritmo da divisão, r_j e q_j tais que $r_{j-2} = r_{j-1}q_j + r_j$ e $0 \leq r_j < r_{j-1}$. Isolando r_j nesta última equação, obtemos

$$r_j = r_{j-2} - r_{j-1}q_j. \tag{4.1}$$

Mas

$$r_{j-2} = ax_{j-2} + by_{j-2} \qquad \text{e} \qquad r_{j-1} = ax_{j-1} + by_{j-1}.$$

Substituindo estes valores em 4.1 obtemos

$$\begin{aligned} r_j &= (ax_{j-2} + by_{j-2}) - (ax_{j-1} + by_{j-1})q_j \\ &= a(x_{j-2} - q_jx_{j-1}) + b(y_{j-2} - y_{j-1}) \end{aligned}$$

Segue que,

$$x_j = x_{j-2} - q_jx_{j-1} \qquad \text{e} \qquad y_j = y_{j-2} - q_jy_{j-1}.$$

Observemos que, para calcular x_j e y_j, precisamos apenas dos valores, x_{j-2}, y_{j-2}, x_{j-1} e y_{j-1}, das duas linhas anteriores à linha j, além do quociente q_j. Portanto, para preencher qualquer linha da tabela, basta conhecer apenas as duas linhas anteriores a ela. Computacionalmente, isto é interessante pois gasta pouca memória (pouco espaço de armazenamento de dados). Vejamos, agora, como iniciar o processo. Para calcular os valores r_1, q_1, x_1 e y_1, precisamos conhecer os valores das duas linhas anteriores. Quanto a r_1 e q_1 não há problema; basta fazer a divisão de a por b obtendo $a = bq_1 + r_1$. Mas, para obter x_1 e y_1 precisamos de x_{-1}, y_{-1}, x_0 e y_0. Observando a tabela 4.1.4, concluímos que devemos ter também

$$a = ax_{-1} + by_{-1} \quad \text{e} \quad b = ax_0 + by_0.$$

Os valores $x_{-1} = 1$, $y_{-1} = 0$, $x_0 = 0$ e $y_0 = 1$ satisfazem estas condições e, portanto, os escolhemos. Assim, com as condições iniciais acima, obtemos, no final do processo, $mdc(a, b) = r_n$, $x = x_n$ e $y = y_n$, onde $mdc(a, b) = ax + by$.

Exemplo 4.1.3.1 Para $a = 2405$ e $b = 630$, obtemos a tabela:

restos	quocientes	x	y
2405	$*$	1	0
630	$*$	0	1
515	3	$1-3\cdot 0=1$	$0-3\cdot 1=-3$
115	1	$0-1\cdot 1=-1$	$1-1\cdot(-3)=4$
55	4	$1-4\cdot(-1)=5$	$-3-4\cdot 4=-19$
$\boxed{5}$	2	$-1-2\cdot 5=\boxed{-11}$	$4-2\cdot(-19)=\boxed{42}$

Portanto,

$$mdc(2405, 630) = 5, x = -11, y = 42 \quad \text{e}$$
$$5 = 2405(-11) + 630(42).$$

Vimos que o algoritmo euclidiano estendido calcula $mdc(a, b)$, bem como um par de inteiros x e y tais que $ax + by = d$. Mas os valores de x e y não são únicos. Na verdade, existe uma infinidade de pares (x, y) de números inteiros que satisfazem esta equação. Por exemplo, se k é um inteiro qualquer e $ax + by = d$, então $(x + kb)a + (y - ka)b = d$. Temos, assim, uma família de pares $(x + kb, y - ka)$, $k \in \mathbb{Z}$ que satisfazem a equação $ax + by = d$.

Para finalizar apresentamos o algoritmo. As variáveis que aparecem no algoritmo são as da tabela seguinte.

restos	quocientes	x	y
.....			
r_1	$*$	x_1	y_1
r_2	$*$	x_2	y_2
r	q	x	y
.....			

Algoritmo 4.1.3.1 (Algoritmo euclidiano estendido) Dados inteiros a e b não simultaneamente nulos e tais que $a \geq b \geq 0$, este algoritmo determina o $mdc(a, b) = d$ e números inteiros x e y tais que $ax + by = d$.

1. se $b = 0$, então $d \leftarrow a$, $x \leftarrow 1$, $y \leftarrow 0$
2. $x_1 \leftarrow 1$, $x_2 \leftarrow 0$, $y_1 \leftarrow 0$, $y_2 \leftarrow 1$
3. enquanto $b > 0$ faça
   - 3.1 $q \leftarrow \lfloor \frac{r_1}{r_2} \rfloor$, $r \leftarrow r_1 - r_2 q$, $x \leftarrow x_1 - x_2 q$, $y \leftarrow y_1 - y_2 q$
   - 3.2 $r_1 \leftarrow r_2$, $r_2 \leftarrow r$, $x_1 \leftarrow x_2$, $x_2 \leftarrow x$, $y_1 \leftarrow y_2$, $y_2 \leftarrow y$
4. $d \leftarrow r_1$, $x \leftarrow x_1$, $y \leftarrow y_1$
5. retorne (d, x, y)

4.2 Método p-1 de Pollard

O método de Pollard é utilizado para encontrar fatores primos p de um número composto n com a propriedade de $p-1$ não ter fatores primos grandes. O método de Pollard não funciona bem para todos os inteiros n, mas quando ele funciona ele é muito eficiente. A idéia na qual o método de Pollard se baseia é dada pela proposição abaixo.

Proposição 4.2.0.1 Sejam n um número inteiro positivo ímpar composto e p um fator primo de n. Sejam a e k números inteiros tais que $mdc(a,p)=1$ e $p-1 \mid k$. Então, $p \mid mdc(a^k-1,n)$.

Prova: Como $p-1 \mid k$, temos que $k=k'(p-1)$, para algum inteiro k'. Como p é primo e $p \nmid a$, segue do pequeno teorema de Fermat que $a^{p-1} \equiv 1 \pmod{p}$. Elevando à k' ambos os termos da congruência e usando a relação entre k e k' obtemos $a^k \equiv 1 \pmod{p}$, que é equivalente a $p \mid a^k-1$. Assim, p é fator comum de a^k-1 e n, donde segue que $p \mid mdc(a^k-1,n)$. □

O método de Pollard, usando a idéia acima, é apresentado a seguir. Tentaremos encontrar um fator primo p de um inteiro composto n. Para tanto, escolhemos inteiros positivos a e k de modo que $mdc(a,n)=1$ (e portanto $mdc(a,p)=1$ para todo fator primo p de n) e k seja divisível por potências de primos pequenos (por exemplo $k=mmc(1,2,3,\ldots,B)$ ou $k=B!$, para um certo inteiro B). Em seguida, calculamos $d=mdc(a^k-1,n)$ e esperamos encontrar um fator não trivial de n (proposição 4.2.0.1). Observemos que não há necessidade de calcular a^k-1; basta calcular $(a^k-1) \bmod n$, pois como vimos na seção 2, $mdc(a,b)=mdc(b,a \bmod b)$. Uma vez calculado d, temos 3 possibilidades.

1. $1<d<n$. Neste caso, encontramos um fator não trivial de n.
2. $d=1$. Este caso ocorre quando $p-1$ não divide k. Devemos então aumentar o valor de k e repetir o processo.
3. $d=n$. O que devemos fazer neste caso é tomar outro valor para a e tentar de novo.

Apresentamos, a seguir, um algoritmo baseado no método de Pollard.

Algoritmo 4.2.0.2 (Algoritmo p-1 de Pollard) Seja $n \geq 2$ um inteiro composto para o qual desejamos achar um fator primo.

1. Escolha um número k que é um produto de primos pequenos elevados a potências pequenas. Por exemplo, considere

$$k = mmc(2, 3, \ldots, B)$$

para algum inteiro B.

2. Escolha um inteiro qualquer a tal que $0 < a < n$.

3. Calcule $mdc(a, n)$. Se ele é estritamente maior que 1, então ele é um fator não trivial de n. Pare. Caso contrário vá para [4].

4. Calcule $d = mdc(a^k - 1, n)$. Se $1 < d < n$, então d é um fator não trivial de n. Pare. Se $d = 1$, vá para [1] e tome um k maior. Se $d = n$, volte para [2] e escolha outro a.

Observe que o algoritmo de Pollard certamente irá parar, pois em algum momento, teremos no passo 1 que $B = \frac{1}{2}(p - 1)$ para algum primo p que divide n, e portanto certamente dividirá k. No entanto, isto gasta muito tempo e o algoritmo não será prático para valores grandes de k. O algoritmo só roda numa quantidade de tempo razoável quando n tem um divisor primo p tal que $p-1$ é produto de primos pequenos elevados a potências pequenas.

Agora que já apresentamos o algoritmo de Pollard vamos mostrar como ele funciona na prática. Tentaremos fatorar o número $n = 275.691.263$. Primeiramente, verificamos que n é composto. Usando o algoritmo exponenciação modular, calculamos $2^{n-1} \mod n = 109.137.477 \neq 1$. Segue do Pequeno Teorema de Fermat que n é composto. Vamos, agora, tentar encontrar um fator primo de n. Tomemos $a = 2$ e $k = mmc(2, 3, 4, 5, 6, 7) = 420$. Em seguida, escrevemos 420 na base 2, obtendo

$$420 = 2^8 + 2^7 + 2^5 + 2^2 = (110100100)_2$$

Calculamos, então, os valores $2^{2^i} \pmod{n}, 0 \leq i \leq 8$, e apresentamo-los na tabela a seguir:

i	$2^{2^i} \mod 275691263$
0	2
1	4
2	16
3	256
4	65536
5	159598351
6	105157287
7	153951677
8	244623874

Usando esta tabela calculamos

$$2^{420} = 2^{2^8+2^7+2^5+2^2} = 2^{2^8} \cdot 2^{2^7} \cdot 2^{2^5} \cdot 2^{2^2} \equiv 252096064 \pmod{275691263}.$$

Um cálculo rápido, usando o algoritmo euclidiano nos dá

$$mdc(2^{420} - 1, n) = mdc(252096063, 275691263) = 1.$$

Assim, vemos que o teste falha. Isto acontece porque n não tem fatores primos p tais que $p-1$ divide 420. O que fazemos então é escolher um valor maior para k esperando que para este novo valor de k, exista um fator primo p de n tal que $p-1|k$. Fazemos $k = mmc(2, 3, 4, 5, \ldots, 13) = 360360$. Como

$$360360 = 2^{18} + 2^{16} + 2^{14} + 2^{13} + 2^{12} + 2^{11} + 2^{10} + 2^9 + 2^8 + 2^7 + 2^5 + 2^3 =$$
$$= (1010111111110101000)_2$$

estendemos a nossa tabela obtendo

i	$2^{2^i} \mod 275691263$
9	229312419
10	177054051
11	257064401
12	52260177
13	252356034
14	211602939
15	49508171
16	182776282
17	89704053
18	89495617

Usando esta tabela obtemos

$$2^{360360} = 2^{2^{18}+2^{16}+2^{14}+2^{13}+2^{12}+2^{11}+2^{10}+2^9+2^8+2^7+2^5+2^3} =$$

$$= 2^{2^{18}} \cdot 2^{2^{16}} \cdot 2^{2^{14}} \cdot 2^{2^{13}} \cdot 2^{2^{12}} \cdot 2^{2^{11}} \cdot 2^{2^{10}} \cdot 2^{2^9} \cdot 2^{2^8} \cdot 2^{2^7} \cdot 2^{2^5} \cdot 2^{2^3} \equiv$$

$$\equiv 197507421 \pmod{275691263}.$$

Agora, usando o algoritmo euclidiano, obtemos

$$mdc(2^{360360} - 1, n) = mdc(197507421, 275691263) = 6553.$$

Encontramos, assim, um fator não trivial de n, como queríamos. Mais precisamente fatoramos n como $n = 6553 \cdot 42071$. Além disso, é fácil verificar que cada um destes fatores é primo, e assim fatoramos n completamente. Tivemos sucesso ao encontrar um fator não trivial de n pois o fator $p = 6553$ encontrado é tal que $p - 1 = 6552 = 2^3.3^2.7.13$ é um fator de $k = 360360 = 2^3.3^2.5.7.11.13$.

É claro que não estamos sugerindo que o algoritmo de Pollard seja necessário para fatorar um número como 275.691.263, pois mesmo um micro computador pode, em poucos segundos, verificar todos os seus possíveis divisores usando o algoritmo trivial. Mas este exemplo revela todas as características do algoritmo que apresentamos.

4.3 Método das Curvas Elíticas

Vamos descrever um método de fatoração de inteiros, devido a H. W. Lenstra, que utiliza curvas elíticas. Ele é baseado no método $p - 1$ de Pollard.

Sejam n um número que sabemos ser composto e p um fator primo de n. O algoritmo de Pollard baseia-se no fato de que os elementos não nulos de $\mathbb{Z}_p$ formam um grupo $\mathbb{Z}_p^*$ de ordem $p - 1$ e que se k é um número inteiro tal que $(p - 1) \mid k$, então $a^k = 1$ no grupo, qualquer que seja $a \in \mathbb{Z}_p^*$. O sucesso na fatoração de n depende do número $p-1$, que é a ordem de $\mathbb{Z}_p^*$, ter decomposição em fatores primos totalmente formada por primos pequenos. No método de Lenstra, substituímos o grupo $\mathbb{Z}_p^*$ por uma família de grupos de ordens diferentes, a saber a família dos grupos $C(\mathbb{Z}_p)$ dos pontos de curvas elíticas sobre $\mathbb{Z}_p$. Assim, ao invés de depositarmos nossas esperanças de sucesso em um único número, $\#\mathbb{Z}_p^* = p - 1$, contamos com a família de números $\#C(\mathbb{Z}_p)$, onde $C(\mathbb{Z}_p)$ é uma curva elítica sobre $\mathbb{Z}_p$. Portanto, se ao tentar fatorar um número composto n, usando o algoritmo de Pollard, não tivermos sucesso, então não teremos outra opção a não ser desistir. Se, contudo, usarmos o método de Lenstra, e ele não funcionar para uma dada

curva, podemos trocá-la por outra e outra, tendo assim mais chances de sucesso.

A tabela a seguir mostra a correspondência entre as notações das operações de grupo usadas nos métodos de Pollard e ECM.

Grupo	$(\mathbb{Z}_p^*)$	$C(\mathbb{Z}_p)$
Operação	Multiplicação módulo p	Adição de pontos
Notação	a e b	P e Q
	Multiplicação: $a \cdot b$	Adição: $P + Q$
	Inverso: a^{-1}	$-P$
	Divisão: a/b	Subtração: $P - Q$
	Exponenciação: a^k	Múltiplo: kP

A idéia na qual se baseia o método de Lenstra é apresentada na seguinte proposição.

Proposição 4.3.0.2 Sejam n um número inteiro positivo composto e p um fator primo de n. Sejam, ainda, $C(\mathbb{Q})$ uma curva elítica sobre $\mathbb{Q}$ dada na forma normal de Weierstrass, P um ponto de $C(\mathbb{Q})$, com $P \neq \mathcal{O} = (0 : 1 : 0)$ e k um inteiro positivo tal que $\#\widetilde{C}(\mathbb{Z}_p) \mid k$. Se $kP = (a : b : c)$, $c \neq 0$, onde $(a : b : c)$ é uma terna de coordenadas normalizadas, então $p \mid mdc(c, n)$.

Prova: Segue do teorema de Lagrange (Proposição 2.1.0.3) que a ordem de qualquer ponto de $\widetilde{C}(\mathbb{Z}_p)$ divide $\#\widetilde{C}(\mathbb{Z}_p)$. Como $\#\widetilde{C}(\mathbb{Z}_p) \mid k$, temos, por transitividade, que a ordem de qualquer ponto de $\widetilde{C}(\mathbb{Z}_p)$ divide k. Em particular, temos $k\widetilde{P} = \widetilde{\mathcal{O}}$. Como a redução módulo p é um homomorfismo de grupos, temos, também, que $\widetilde{kP} = k\widetilde{P}$, donde segue que $\widetilde{kP} = \widetilde{\mathcal{O}}$, isto é, a redução módulo p do ponto kP é o ponto no infinito $\widetilde{\mathcal{O}}$ em $\mathbb{P}^2(\mathbb{Z}_p)$. Se $kP = \left(\frac{a}{c} : \frac{b}{c} : 1\right) = (a : b : c)$, onde $(a : b : c)$ é uma terna de coordenadas normalizadas, então $\tilde{c} = 0$, onde 0 é o elemento neutro de $\mathbb{Z}_p$. Logo, $c \equiv 0 \pmod{p}$, donde segue que $p \mid c$. Como $p \mid n$, obtemos a relação $p \mid mdc(c, n)$, como queríamos. □

Antes de apresentar o método de Lenstra, faremos algumas considerações a respeito do cálculo de kP, que é uma etapa importante do método. Dados um inteiro k e uma curva elítica $C(\mathbb{Q})$, queremos calcular kP eficientemente. Para tanto, primeiramente, expressamos k em termos da sua expansão binária, isto é, escrevemos

$$k = k_0 + k_1 \cdot 2 + k_2 \cdot 2^2 + ... + k_r \cdot 2^r,$$

onde $k_i = 0$ ou $k_i = 1$, $1 \leq i \leq r$ e $r \leq \log_2 k$. Calculamos, em seguida, os pontos $2^i P$, $1 \leq i \leq r$, e obtemos

$$\begin{aligned}
&P_0 = P \\
&P_1 = 2P_0 = 2P \\
&P_2 = 2P_1 = 2^2 P \\
&P_3 = 2P_2 = 2^3 P \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&P_{r-1} = 2P_{r-2} = 2^{r-1} P \\
&P_r = 2P_{r-1} = 2^r P
\end{aligned}$$

Para obter kP, somamos todos os pontos $P_i's$ para os quais $k_i = 1$.

$$kP = \sum k_i(2^i P), \quad k_i = 1$$

Dessa maneira, calculamos kP com um número de passos menor do que $2\log_2 k$, onde cada passo é uma adição ou uma duplicação de pontos.

Para obter as coordenadas de kP no método de Lenstra, não vamos fazer os cálculos considerando coordenadas racionais, pois os numeradores e os denominadores teriam aproximadamente k^2 dígitos, levando um tempo muito grande. Por outro lado, não podemos fazer as operações módulo p, pois não sabemos quem é p. Faremos, então, as operações módulo n. Como n não é primo, $\mathbb{Z}_n$ não é um corpo, mas sim um anel. Sendo assim, alguns elementos $c \in \mathbb{Z}_n$ não são invertíveis, isto é, $mdc(c, n) \neq 1$. Lembremos que um elemento $c \in \mathbb{Z}_n$ é invertível se, e somente se, $mdc(c, n) = 1$.

Para somar dois pontos $P_1 = (x_1, y_1)$ e $P_2 = (x_2, y_2)$, usamos as fórmulas

$$\lambda = \frac{y_2 - y_1}{x_2 - x_1} = (y_2 - y_1) \cdot (x_2 - x_1)^{-1},$$

$$x_3 = \lambda^2 - x_1 - x_2,$$

$$y_3 = \lambda(x_1 - x_3) - y_1.$$

Se $x_2 - x_1$ for invertível em $\mathbb{Z}_n$, isto é, $mdc(x_2 - x_1, n) = 1$, então calculamos seu inverso utilizando o algoritmo euclidiano estendido e obtemos $(P_1 + P_2) \mod n$; caso contrário, não poderemos calcular $(P_1 + P_2) \mod n$. Para duplicar um ponto $P_1 = (x_1, y_1)$, usamos as fórmulas

$$\lambda = \frac{3x_1^2+a}{2y_1} = (3x_1^2 + a) \cdot (2y_1)^{-1},$$

$$x_3 = \lambda^2 - 2x_1,$$

$$y_3 = \lambda(x_1 - x_3) - y_1.$$

Se $2y_1$ for invertível em $\mathbb{Z}_n$, isto é, $mdc(2y_1, n) = 1$, calculamos seu inverso utilizando o algoritmo euclidiano estendido e obtemos $(2P_1) \bmod n$; caso contrário, não poderemos calcular $(2P_1) \bmod n$.

Então, na verdade, não temos operações em $C(\mathbb{Z}_n)$, mas pseudo operações. Poderemos calcular a soma ou a duplicação de pontos desde que o denominador de λ seja invertível em $\mathbb{Z}_n$.

Feitas estas considerações apresentamos, agora, o método de Lenstra. Dado um inteiro n que sabemos ser composto, vamos tentar encontrar um fator primo p de n. Escolhemos um ponto racional $P = (x_1, y_1)$, uma curva elítica que contém P e um número inteiro k que é o produto de primos pequenos elevados à potências pequenas (e.g., $k = mmc(1, 2, \ldots, B)$ ou $k = B!$, onde B é um certo inteiro dado). Em seguida, tentamos calcular kP pelo método descrito acima. Temos duas possibilidades:

1. Conseguimos calcular kP. Então, cada λ que surgiu no cálculo de kP têm denominador c invertível em $\mathbb{Z}_n$, e portanto $mdc(c, n) = 1$. Nesse caso, não obteremos um fator não trivial de n. Devemos, então, aumentar o valor de k ou tomar uma outra curva e tentar de novo.

2. Não conseguimos calcular kP. Então, em alguma etapa do processo, nos deparamos com um λ cujo denominador não é invertível em $\mathbb{Z}_n$. Nesse caso, se c é o denominador desse λ, então $1 < mdc(c, n) \leq n$, isto é, $mdc(c, n)$ é um fator não trivial ou um múltiplo de n. Se $mdc(c, n) < n$ conseguimos fatorar n; se $mdc(c, n) = n$ não tivemos sorte; tomamos uma outra curva e repetimos o processo.

Observemos que o método de Lenstra nos dá um fator não trivial de n exatamente quando a adição ou a duplicação falha. Mostramos, assim, a essência do método. É claro que na prática existem várias melhorias que o tornam mais eficiente.

A seguir, apresentamos o Algoritmo de Lenstra.

Algoritmo 4.3.0.3 (Algoritmo de Lenstra de fatoração de inteiros) Seja $n \geq 2$ um inteiro composto para o qual desejamos achar um fator primo.

1. Verifique que $mdc(n, 6) = 1$ e que n não tem a forma m^r para algum $r \geq 2$.

2. Escolha inteiros aleatórios a, x_1 e y_1 entre 1 e n.

3. Faça $b = y_1^2 - x_1^3 - ax_1 \pmod{n}$ (Seja C a curva $y^2 = x^3 + ax + b$ e $P = (x_1, y_1)$ um ponto de C).

4. Verifique que $mdc(4a^3 + 27b^2, n) = 1$. (Se for igual a n, vá para passo [2] e escolha novo a. Se estiver entre 1 e n, então ele é um fator não trivial de n. Pare.)

5. Escolha um número k que é um produto de primos pequenos elevados a potências pequenas. Por exemplo, considere

$$k = mmc(1, 2, 3, \ldots, B)$$

   para algum inteiro B.

6. Tente calcular $kP \bmod n$ pelo método das duplicações sucessivas. Se conseguir (é porque todos os λ's têm denominadores invertíveis em $\mathbb{Z}_n$), vá para [5] e aumente o valor de k ou vá para [2] e tome outra curva. Caso contrário (é porque em alguma etapa do cálculo de kP o denominador c de λ é não invertível, i.e., $mdc(c, n) \neq 1$); vá para [7].

7. Se $mdc(c, n) < n$, encontramos um fator não trivial de n. Pare. Se $mdc(c, n) = n$ vá para [2] e escolha outra curva.

Para uma melhor compreensão de como o algoritmo funciona, vamos utilizá-lo para fatorar o número $n = 5707444801$. Primeiramente, verificamos que n é composto aplicando o pequeno teorema de Fermat. Fazemos isto, calculando $2^{n-1} \mod n = 2^{5707444800} \mod 5707444801 = 4650752551 \neq 1$. Tendo mostrado que n é composto, verificamos, facilmente, que $mdc(n, 6) = 1$ e que ele não é uma potência de inteiros, o que pode ser feito mesmo com uma calculadora. Além disso, n tem um fator primo p menor que $\lfloor\sqrt{5707444801}\rfloor = 75547$. Escolhemos, então, $k = mmc(1, 2, 3, \ldots, 16, 17) = 12252240$, que é formado por potências de primos pequenos, e esperamos que um inteiro próximo de p divida k ($p + 1 - 2\sqrt{p} \leq \#C(\mathbb{Z}_p) \leq p + 1 + 2\sqrt{p}$). Em seguida, escolhemos um ponto P e uma curva elítica $y^2 = x^3 + ax + b$ que contém P. Fazendo $P = (-1, 1)$ e substituindo as coordenadas de P na equação da curva, obtemos a relação $b = 2 + a$ entre a e b. Assim, para cada valor de a, encontramos um valor correspondente para b, obtendo assim uma família de curvas que contém o ponto $P = (-1, 1)$. Para $a = 1$, obtemos que $b = 3$, isto é, escolhemos a curva $C : y^2 = x^3 + x + 3$ que

contém o ponto $P(-1,1)$. O próximo passo é calcular $kP \mod n$ usando o método das duplicações sucessivas. Para tanto, escrevemos, primeiramente a expansão binária de k,

$$k = 12252240 = 2^4+2^6+2^{10}+2^{12}+2^{13}+2^{14}+2^{15}+2^{17}+2^{19}+2^{20}+2^{21}+2^{23}.$$

Depois, calculamos, 2^iP, $1 \leq i \leq 23$. Veja tabela a seguir.

i	$2^iP \pmod{5707444801}$
0	$(-1,1)$
1	$(6, 5707444786)$
2	$(754651036, 839417157)$
3	$(2222582142, 95809934)$
4	$(5049127219, 4599744024)$
5	$(4807555263, 4375841951)$
6	$(4565085608, 958930015)$
7	$(5282632356, 2147896970)$
8	$(2533266988, 2260001774)$
9	$(1880890231, 2542734214)$
10	$(3784280113, 4940296468)$
11	$(2896346268, 1996850782)$
12	$(3043898975, 580042499)$
13	$(1566382399, 3248476863)$
14	$(5630555204, 1639313199)$
15	$(1189600119, 1274415959)$
16	$(1561165783, 3927311419)$
17	$(3785615482, 3141397759)$
18	$(1998998795, 2701731382)$
19	$(3977114934, 2436820414)$
20	$(959556823, 1333239783)$
21	$(2513598784, 4392088587)$
22	$(247308551, 2550492208)$
23	$(2867105003, 2094327781)$

Em seguida, calculamos as somas parciais dos valores $2^iP \pmod n$ considerando aqueles $i's$ que aparecem na expansão binária de k. Apresentamos tais valores na tabela seguinte.

$$
\begin{array}{rcc}
2^4P & = & 16P = (2363120125, 5125661596) \\
2^4P + 2^6P & = & 80P = (1987517093, 1487435672) \\
2^4P + 2^6P + 2^{10}P & = & 1104P = (2928961367, 2134284297) \\
2^4P + 2^6P + 2^{10}P + 2^{12} & = & 5200P = (1782311891, 2836630032) \\
\text{(soma parcial anterior)} + 2^{13}P & = & 13392P = (3907238149, 181856945) \\
\text{(soma parcial anterior)} + 2^{14}P & = & 29776P = (946309593, 12996932) \\
\text{(soma parcial anterior)} + 2^{15}P & = & 62544P = (3129416259, 653861891) \\
\text{(soma parcial anterior)} + 2^{17}P & = & 193616P = (3959279191, 2728642856) \\
\text{(soma parcial anterior)} + 2^{19}P & = & 717904P = (2013008712, 3316614920) \\
\text{(soma parcial anterior)} + 2^{20}P & = & 1766480 = (5558287732, 1878021496) \\
\text{(soma parcial anterior)} + 2^{21}P & = & 3863632P = (3878344933, 2261189212) \\
\text{(soma parcial anterior)} + 2^{23}P & = & 12252240P = (3691148282, 2533039544)
\end{array}
$$

Encontramos, assim, o ponto

$$kP \mod n = 12252240P \mod 5707444801 = (3691148282, 2533039544)$$

sobre a curva $C : y^2 = x^3 + x + 3$. O fato de termos conseguido calcular kP significa que as operações de adição e duplicação necessárias foram todas possíveis. Como mencionamos anteriormente, quando isso acontece não obtemos fator de n. Devemos, então, como está descrito no algoritmo, aumentar o valor de k ou usar uma nova curva. Escolhemos esta última opção. Mantemos, então, $k = 12252240$, $P = (-1, 1)$ e tomamos um novo valor para a a saber $a = 2$, e assim $b = 2 + a = 4$. Usando esta nova curva $C : y^2 = x^3 + 2x + 4$ e repetindo os cálculos acima, conseguimos ainda obter $kP \bmod n$. Continuando a trocar curvas, fazendo $a = 3, 4, 5 \ldots, 11$, somos ainda capazes de calcular kP. No entanto, quando fazemos $a = 12$ e $b = 14$, a lei da duplicação falha e achamos um fator não trivial de n. Conseguimos calcular $2P$ e $2^2P = 4P$. Mas não foi possível calcular $2^3P = 8P$, pois o número 87045100 não é invertível em $\mathbb{Z}_n$ e, portanto, não é possível calcular $\lambda = \frac{3(3919984703)^2+12}{2 \cdot 87045100} \mod n$. Veja tabela a seguir.

$P = (-1, 1)$ e $y^2 = x^3 + 12x + 14$

i	$2^iP \mod (5707444801)$
0	$(-1, 1)$
1	$(4280583659, 2140291355)$
2	$(3919984703, 87045100)$

Encontramos um fator não trivial de n, calculando $mdc(87045100, n) = 51203$, obtendo, assim, a fatoração $5707444801 = 111467 \times 51203$. Como ambos os fatores são números primos, obtemos a fatoração completa de n.

Consideremos, agora, o exemplo de um número inteiro para o qual a obtenção de um fator primo deve-se à falha na operação de adição de pontos.

Seja $n = 2263295989$ tal número. Como antes, verificamos que n é composto usando o pequeno teorema de Fermat. Calculamos

$$2^{n-1} \mod n = 2^{2263295988} \mod 2263295989 = 87926644 \neq 1.$$

Tendo mostrado que n é composto, verificamos, facilmente, que $mdc(n, 6) = 1$ e que ele não uma potência de inteiros.

Além disso, n tem um fator primo p menor que $\lfloor\sqrt{2263295989}\rfloor = 47574$. Escolhemos, de novo, $k = mmc(1, 2, 3, \ldots, 16, 17) = 12252240$, que é formado por potências de primos pequenos, e esperamos que um inteiro próximo de p divida k ($p + 1 - 2\sqrt{p} \leq \#C(\mathbb{Z}_p) \leq p + 1 + 2\sqrt{p}$). Em seguida, escolhemos um ponto P e uma curva elítica $y^2 = x^3 + ax + b$ que contém P. Fazendo $P(1, 1)$ e substituindo as coordenadas de P na equação da curva obtemos a relação $b = -a$ entre a e b. Assim, para cada valor de a obtemos um valor correspondente para b, obtendo assim uma família de curvas que contém o ponto $P = (1, 1)$.

Como no exemplo anterior, testamos vários valores para a. Para todos os valores de a entre 1 e 16, conseguimos calcular $kP \mod n$. Considerando $a = 17$ ($b = -17$), obtemos a tabela a seguir, apresentando os pontos da forma 2^iP sobre a curva $y^2 = x^3 + 17x - 17$.

i	$2^iP \pmod{2263295989}$
0	$(1, 1)$
1	$(98, 2263295018)$
2	$(2114238360, 2046943107)$
3	$(1294928621, 345290563)$
4	$(2183586607, 591653227)$
5	$(374988228, 2159455512)$
6	$(1120409521, 1833989734)$
7	$(1269278042, 1293551569)$
8	$(1403534318, 2008424492)$
9	$(955316660, 1964856163)$
10	$(110844640, 407314189)$
11	$(1652815329, 958428251)$
12	$(1441240225, 1402896651)$
13	$(996669408, 312124004)$
14	$(2215448523, 616101746)$
15	$(1643275943, 498901537)$
16	$(368583260, 943302883)$
17	$(1148615570, 1677239360)$
18	$(1465937153, 466724747)$
19	$(1937811695, 1730806812)$
20	$(1580816434, 1834323249)$
21	$(616227550, 28913960)$
22	$(369273563, 503263872)$
23	$(547396470, 1241765356)$

Em seguida, calculamos as somas parciais dos valores $2^iP \bmod n$ considerando aqueles $i's$ que aparecem na expansão binária de k, com exceção do valor $i = 23$. Apresentamos tais valores na tabela seguinte.

$$
\begin{array}{rcc}
2^4P & = & 16P = (432964500, 1611749369) \\
2^4P + 2^6P & = & 80P = (1713379813, 2163156340) \\
2^4P + 2^6P + 2^{10}P & = & 1104P = (4072568, 2060677994) \\
2^4P + 2^6P + 2^{10}P + 2^{12} & = & 5200P = (312777565, 2199051011) \\
\text{(soma parcial anterior)} + 2^{13}P & = & 13392P = (263480901, 1020096372) \\
\text{(soma parcial anterior)} + 2^{14}P & = & 29776P = (181478733, 3711027) \\
\text{(soma parcial anterior)} + 2^{15}P & = & 62544P = (1731356989, 1724846512) \\
\text{(soma parcial anterior)} + 2^{17}P & = & 193616P = (2237442378, 484451154) \\
\text{(soma parcial anterior)} + 2^{19}P & = & 717904P = (1813948455, 1077057578) \\
\text{(soma parcial anterior)} + 2^{20}P & = & 1766480 = (445505156, 284697114) \\
\text{(soma parcial anterior)} + 2^{21}P & = & 3863632P = (149249255, 2095738440)
\end{array}
$$

Quando tentamos calcular as somas parciais para obter $kP \mod n$, obtemos, no penúltimo passo, a expressão

$$(2^4 + 2^6 + \ldots + 2^{20} + 2^{21})P = 3863632P = (149249255, 2095738440).$$

Para obter kP, temos que somar este ponto com o ponto

$$2^{23}P = (547396470, 1241765356) \pmod{n},$$

i.e., calcular

$$kP = (149249255, 2095738440) + (547396470, 12417653561) \pmod{n}.$$

Para tanto, precisamos calcular o inverso módulo n da diferença das coordenadas x. Mas tal inverso não existe pois

$$mdc(149249255 - 547396470, n) = 51407 \neq 1.$$

Assim, o cálculo de $12252240(1, 1)$ sobre a curva $C : y^2 = x^3 + 16x - 17$ falha, mas nos leva à fatoração $n = 2263295989 = 51407 \times 44027$. Como ambos os fatores são números primos, obtemos a fatoração completa de n.

Capítulo 5

Conclusão

Tendo escolhido o método da curvas elíticas para estudarmos, pensamos, inicialmente, em vários caminhos a seguir: a descrição do método, a sua análise, o estudo de métodos para escolher as curvas apropriadas, a implementação de um algoritmo de fatoração e a comparação com outros métodos de fatoração. Ao estudarmos o artigo de Lenstra [Len87] e pesquisar sobre a literatura existente, optamos pela descrição e a fundamentação matemática do Método das Curvas Elíticas, acreditando que um trabalho nesta linha possa contribuir para que estudantes da área de computação se interessem por este assunto e se sintam mais confortáveis para entender a matemática envolvida.

Referências Bibliográficas

[AKS02] M. Agrawal, N. Kayal, e N. Saxena. Primes is in P, Agosto 2002. IIT Kanpur - http://www.cse.iitk.ac.in/news/primality.html.

[AM93] A. O. L. Atkin e F. Morain. Find suitable curves for the elliptic curve method of factorization. *Math. Comp.*, 60:399–405, 1993.

[Bre89] D. M. Bressoud. *Factorization and Primality Testing*. Springer-Verlag, New York, 1989.

[Bre00] R. P. Brent. Recent progress and prospects for integer factorisation algorithms, Julho 2000. Lecture Notes in Computer Science, Vol. 1858, Springer-Verlag, Berlin, 2000, 3-22. Preliminary version available in ftp://ftp.comlab.ox.ac.uk/pub/Documents/ techpapers/ Richard.Brent/rpb196tr.ps.gz.

[Buc02] J. A. Buchmann. *Introdução à Criptografia, Tradução de Bázan Tecnologia e Lingüística*. Berkeley Brasil, 2002.

[CLR90] T. H. Cormen, C. E. Leiserson, e R. L. Rivest. *Introduction to Algorithms*. MIT Press/McGraw-Hill, 1990.

[Coh93] H. Cohen. *A course in computational algebraic number theory*. Graduate texts in Math. **138**, Springer-Verlag, Heidelberg, 1993.

[Cou97] S. C. Coutinho. *Números inteiros e criptografia RSA*. Série de Computação e Matemática, IMPA/SBM, Rio de Janeiro, 1997.

[Ful69] W. Fulton. *Algebraic curves*. Benjamin, 1969.

[GL02] A. Garcia e Y. Lequain. *Elementos de Álgebra Abstrata*. IMPA, 2002.

[Gon79] A. Gonçalves. *Introdução à Álgebra*. IMPA, 1979.

[Hef93] A. Hefez. *Curso de Álgebra, Volume 1*. IMPA, 1993.

[Knu81] D. E. Knuth. *The Art of Computer Programming, Vol. 2: Seminumerical Algorithms*. Addison-Wesley, Reading, MA, 1981.

[Kob87a] N. Koblitz. *A course in number theory and cryptography. Graduate texts in Math. 114*. Springer-Verlag, New York, 1987.

[Kob87b] N. Koblitz. Elliptic curve cryptosystems. *Mathematics of Computation*, 48:203–209, 1987.

[LD98] J. López e R. Dahab. On computing a multiple of an elliptic curve point, Abril 1998. Relatório Técnico, IMECC-UNICAMP-98-13.

[Len87] H. W. Lenstra. Factoring integers with elliptic curves. *Annals of Math.*, 126:649–673, 1987.

[LLMP89] A. K. Lenstra, H. W. Lenstra, M. S. Manasse, e J. M. Pollard. *The Number Field Sieve, in The development of the number field sieve. pp 11–42, Lecture Notes in Mathematics 1554*. Springer-Verlag, 1989.

[Luc86] C. L. Lucchesi. *Introdução à Criptografia Computacional*. Editora da UNICAMP/Editora Papirus, Campinas, 1986.

[MC00] C. P. Milies e S. P. Coelho. *Números, Uma Introdução à Matemàtica*. EDUSP, 2000.

[Mil86] V. S. Miller. Uses of elliptic curves in cryptography. *Advances in Cryptology - CRYPTO'85*, 218:417–426, 1986.

[Mon94] P. L. Montgomery. A survey of modern integer factorization algorithms. *CWI Quarterly*, 7:337–366, 1994.

[MvOV96] A. J. Menezes, P. C. van Oorschot, e S. A. Vanstone. *Handbook of Applied Cryptography*. CRC Press, 1996.

[NZM91] I. Niven, H. S. Zuckerman, e H. L. Montgomery. *An Introduction to the Theory of Numbers*. John Wiley & Sons, fifth edition, 1991.

[Odl95] A. Odlyzko. The future of integer factorization, technical report, rsa laboratories cryptobytes, Julho 1995. Http://www.rsasecurity.com/rsalabs/cryptobytes/index.html.

[Pol74] J. M. Pollard. Theorems on factorization and primality testing. *Proc. Camb. Philo. Soc.*, 76:521–528, 1974.

[RSA78] R. L. Rivest, A. Shamir, e L. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Comm. ACM.*, 21:120–126, 1978.

[Sil86] J. H. Silverman. *The Arithmetic of Elliptic Curves*. Springer-Verlag, New York, 1986.

[ST92] J. H. Silverman e J. Tate. *Rational points on elliptic curves, Undergraduate texts in Math*. Springer-Verlag, New York, 1992.