Universidade Estadual de Campinas
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E COMPUTAÇÃO CIENTÍFICA
Departamento de Matemática

Dissertação de Mestrado

Curvas Elípticas: algumas aplicações em Criptografia e em Teoria dos Números

por

Karina Kfouri Sartori

Orientador: Prof. Dr. Paulo Roberto Brumatti

FICHA CATALOGRÁICA ELABORADA PELA BIBLIOTECA DO IMECC DA UNICAMP
Bibliotecária: Maria Júlia Milani Rodrigues - CRB8a /2116

Sartori, Karina Kfouri
Sa77c Curvas Elípticas: algumas aplicações em Criptografia e em Teoria dos Números / Karina Kfouri Sartori– Campinas, [S.P.:s.n.], 2006.

Orientador: Paulo Roberto Brumatti
Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Cientíica.

1. Curvas Elípticas. 2. Criptografia. 3. Teoria dos Números. I. Brumatti, Paulo Roberto. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Título em inglês: Elliptic curves: some applications in Cryptography and Number Theory.

Palavras-chave em inglês (Keywords): 1. Elliptic Curves 2. Cryptography. 3. Number Theory.

Área de concentração: Álgebra

Titulação: Mestre em Matemática

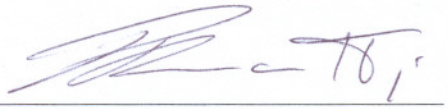
Banca examinadora: Prof. Dr. Paulo Roberto Brumatti (IMECC-UNICAMP)
Prof. Dr. Fernando Eduardo Torres Orihuela (IMECC-UNICAMP)
Profa. Dra. Neuza Kazuko Kakuta (IBILCE-UNESP)

Data da defesa: 12/04/2006

Curvas Elípticas: algumas aplicações em Criptografia e em Teoria dos Números

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Karina Kfouri Sartori e aprovada pela comissão julgadora.

Campinas, 09 de maio de 2006



Prof. Dr. Paulo Roberto Brumatti
Orientador

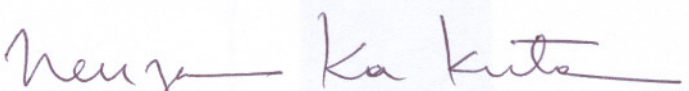
Banca Examinadora:

1. Prof. Dr. Paulo Roberto Brumatti
2. Prof. Dr. Fernando Eduardo Torres Orihuela
3. Profa. Dra. Neuza Kazuko Kakuta

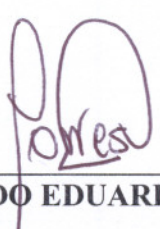
Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica, UNICAMP, como requisito parcial para obtenção do Título de **Mestre em Matemática**.
Área de concentração: **Álgebra**.

Dissertação de Mestrado defendida em 12 de abril de 2006 e aprovada pela Banca Examinadora composta pelos Profs. Drs.

Prof. (a). Dr (a). PAULO ROBERTO BRUMATTI



Prof. (a). Dr (a). NEUZA KAZUKO KAKUTA



Prof. (a). Dr (a). FERNANDO EDUARDO TORRES ORIHUELA

Agradecimentos

Gostaria de fazer os seguintes agradecimentos:

- Ao Prof. Dr. Paulo Roberto Brumatti por sua atenção e orientação na elaboração desta dissertação
- Ao Prof. Dr. Fernando Eduardo Torres Orihuela pelo apoio
- Aos professores dos IBILCE/UNESP, em especial à Profa. Dra. Neuza Kazuko Kakuta pelo apoio e incentivo
- Aos meus amigos e minha família pelo estímulo
- A CNPQ pela ajuda financeira

Resumo

O objetivo central de estudo neste trabalho é introduzir o conceito de curvas elípticas. Tal assunto é clássico dentro da geometria algébrica e tem aplicações em Criptografia e Teoria dos Números. Neste trabalho descrevemos algumas delas: em Criptografia, apresentamos sistemas análogos aos de Diffie-Helman, Massey-Omura e ElGamal que são baseados no grupo abeliano finito de um curva elíptica definida sobre um corpo finito. Em Teoria dos Números descrevemos o método de Lenstra para descobrir fatores primos de um número inteiro, que, por sinal, também tem uma relação muito estreita com certo tipo de sistema criptográfico. Ainda em Teoria dos Números, apresentamos uma caracterização de números congruentes através da estrutura do grupo de uma determinada curva elíptica.

Abstract

The central objective of study in this work is to introduce the concept of elliptic curves. Such subject is classic inside of algebraic geometry and has applications in Cryptography and Number Theory. In this work we describe some of them: in Cryptography, we present analogous systems to the ones of Diffie-Helman, Massey-Omura and ElGamal that are based on the finite abelian group of an elliptic curve defined over a finite field. In Number Theory, we describe the method of Lenstra to discover prime factors of a whole number, that, by the way, also has a very narrow relation with certain type of cryptosystem. Still in Number Theory, we present a characterization of congruentes numbers through the structure of the group of one determined elliptic curve.

Introdução

Um tópico da geometria algébrica - curvas elípticas - além de ter uma literatura muito extensa, tem encontrado aplicações em diversos ramos da matemática. Por exemplo, a pouco tempo elas foram usadas na solução de um dos grandes teoremas da história da Matemática: o Último Teorema de Fermat.

Já a algum tempo, muita atenção tem sido voltada ao uso de curvas elípticas na criptografia com chave pública, proposto primeiramente nos trabalhos de Koblitz [6] e Miller [9]. A motivação para isso é o fato de que além de não se conhecer nenhum algorítmo sub-exponencial para resolver o problema do logarítmo discreto em uma curva elíptica geral, as curvas elípticas sobre corpos finitos nos fornecem uma grande fonte de grupos abelianos finitos. Além disso, como discutiremos neste trabalho, os criptossistemas que empregam o problema do logarítmo discreto em corpos finitos, tais como os sistemas de Diffie-Hellman, Massey-Omura e ElGamal, têm análogos no caso das curvas elípticas.

Além da criptografia, veremos aplicações das curvas elípticas na Teoria dos Números, mais precisamente na fatoração de um número e na caracterização de um número congruente.

Começamos apresentando no primeiro capítulo definições e resultados importantes sobre curvas elípticas, como por exemplo, introduzindo a adição entre pontos de uma curva elíptica e demonstrando que a curva elíptica com tal operação é um grupo abeliano. No fim do capítulo também fazemos alguma considerações importantes a respeito de curvas elípticas sobre corpos finitos, que serão importantes no decorrer do texto.

No segundo capítulo, inicialmente introduzimos os conceitos básicos de criptografia e depois apresentamos exemplos de sistemas criptográficos com chave pública, como é o caso do sistema RSA, baseado no problema de se fatorar um número inteiro e os sistemas de Diffie-Hellman, Massey-Omura e ElGamal, baseados no problema do logarítmo discreto.

Finalmente, no terceiro e último capítulo, utilizaremos os conceitos apresentados nos capítulos anteriores para vermos algumas aplicações das curvas elípticas em Critptografia e em Teoria dos Números. Em Criptografia, tendo como ponto de partida o artigo "Elliptic curve cryptosystems", de N. Koblitz, apresentamos sistemas criptográficos análogos aos de Diffie-Hellman,

Massey-Omura e ElGamal. Já em Teoria dos Números, descrevemos o método de Lenstra para fatorar um número composto e apresentamos uma caracterização de números congruentes, baseado no artigo "Factoring integers with elliptic curves"de H. W. Lenstra e no livro "Introduction to Elliptic Curves and Modular Forms". de N. Koblitz, respectivamente.

Para encerrar gostaríamos de acrescentar que apresentamos nessa dissertação apenas uma parte muito pequena de uma teoria clássica que é muito vasta, atual e rica.

Conteúdo

Capítulo 1

Curvas Elípticas

O objetivo desse capítulo é definir curvas elípticas e demonstrar que as cúbicas não singulares possuem uma estrutura de grupo aditivo. E, como consequência disso, teremos que se C é uma curva elíptica sobre K, então $C(K)$ juntamente com a adição de pontos é um grupo comutativo.

1.1 Preliminares

Nesta seção, introduziremos definições e resultados que serão relevantes para um melhor entendimento das curvas elípticas.

1.1.1 Curva Algébrica Plana

Definição 1.1.1. *Seja K um corpo. Uma curva algébrica plana sobre K é o lugar geométrico dos pontos de K^2 cujas coordenadas cartesianas satisfazem a uma equação do tipo*

$$f(X, Y) = 0,$$

onde f é um polinômio não constante.

Exemplo 1.1.1. *O círculo de raio r e centro (a, b), é o lugar dos pontos que satisfazem a equação*

$$(X - a)^2 + (Y - b)^2 - r^2 = 0$$

□

1.1.2 Equação de uma curva algébrica

Admitiremos nessa seção que o corpo K é algebricamente fechado e sua característica é zero. Neste caso, uma curva algébrica em K^2 fica determinada pelos fatores irredutíveis de uma equação que definimos anteriormente. É o que apresentamos no resultado seguinte.

Proposição 1.1.1. *Sejam $f, g \in K[X,Y]$. Então $f(X,Y) = 0$ e $g(X,Y) = 0$ têm as mesmas soluções em K^2 se, e somente se, os fatores irredutíveis de f,g são os mesmos.*

Demonstração: É evidente que se f e g têm os mesmos fatores irredutíveis então $f(X,Y) = 0 = g(X,Y)$ tem o mesmo conjunto solução.

Reciprocamente, seja $p \in K[X,Y]$ um fator irredutível de f. Por hipótese, para cada $(x,y) \in K^2$, vale a implicação,

$$p(x,y) = 0 \Rightarrow g(x,y) = 0 \tag{1.1}$$

Trocando X por Y se necessário, podemos supor que Y ocorre efetivamente em p, isto é,

$$p = p(x,Y) = a_n(x)Y^n + \ldots + a_1(x)Y + a_0(x);\ n \geq 1,\ a_i(x) \in A,\ \forall i \text{ e } a_n(x) \neq 0 \tag{1.2}$$

Seja $A = K[X]$ e $L = K(X)$ o corpo de frações de A, pelo lema de Gauss, $p \in A[Y]$ é irredutível em $L[Y]$.

Suponhamos, por absurdo, que p não divide g.

Então, $MDC(p,g) = 1$.

Daí, existe uma relação

$$ap + bg = 1,$$

onde $a, b \in L[Y]$.

Podemos, então, escrever

$$a = \frac{a'}{c}, \quad b = \frac{b'}{c}$$

com $a', b' \in A[Y]$ e $c \in A$, $c \neq 0$.

Assim, obtemos

$$a'p + b'g = c.$$

Observando a equação (1.2) podemos ver que, exceto para um número finito de valores de x, o polinômio $p(x, Y)$ é não constante e portanto $p(x, Y) = 0$ admite solução, pois K é um corpo algebricamente fechado.

Conclui-se, então, por (1.1), que há uma infinidade de valores de x tais que $c(x) = 0$, e portanto $c = 0$.

Esta contradição mostra que p divide g em $L[Y]$, seguindo, novamente pelo lema de Gauss, que p divide g em $K[X, Y]$. □

Podemos deduzir, como consequência da proposição anterior, que uma curva algébrica, dada como o lugar das soluções de uma equação polinomial não constante $f(X, Y) = 0$, determina, a menos de fator constante, uma equação de grau mínimo; basta tomar o produto dos fatores irredutíveis distintos de f.

Este fato induz uma substituição da definição 1.1.1 pela seguinte, onde passamos a identificar "curva"com sua equação.

Definição 1.1.2. *Sejam $f, g \in K[X, Y] \setminus K$.*

Definimos a relação,

$$f \sim g \iff \exists\ \lambda \in K^* \ ;\ f = \lambda g.$$

Uma curva algébrica plana afim, ou mais abreviadamente uma curva, definida por $f \in K[X, Y]$ é a classe de equivalência de f.

Sendo K algebricamente fechado, a equação de uma curva é qualquer um dos polinômios nessa classe.

O conjunto $V(f) = \{(x, y) \in K^2 / f(x, y) = 0\}$ é chamado traço da curva f. O grau de uma curva f é o grau total de sua equação.

Notação: Denotaremos a curva f por $f = 0$ e o grau de uma curva f por ∂f.

Observação 1.1.1. *As curvas $X^2 = 0$ e $X = 0$ têm o mesmo traço, mas são curvas distintas.*

Definição 1.1.3. *Uma curva $f = 0$ é irredutível se $f \in K[X, Y]$ é irredutível.*

Exemplo 1.1.2. *As parábolas, elipses e hipérboles são curvas irredutíveis.* □

Como $K[X,Y]$ é um domínio fatorial então para todo $f \notin K[X,Y]^*$, existem $p_1, \ldots, p_r$ irredutíveis em $K[X,Y]$ tais que:

$$f = p_1^{m_1} \ldots p_r^{m_r}, \text{ com } 1 \leq m_i, \ m_i \in \mathbb{N}, i = 1, \ldots, r.$$

Definição 1.1.4. *As curvas $p_i = 0, i = 1, \ldots, r$ são chamadas de componentes irredutíveis das curva $f = 0$, e m_i é chamado de multiplicidade da componente p_i.*

Vamos enunciar agora um teorema clássico (daremos agora a sua versão fraca, a versão mais geral será enunciada mais adiante) que nos dá informações sobre a cardinalidade da interseção de duas curvas. Sua demonstração pode ser vista em [3]

Teorema 1.1.1 (Teorema de Bézout (versão fraca))**.** *Seja K um corpo. Sejam $f(X,Y)$ e $g(X,Y)$ dois polinômios em $K[X,Y]$ de graus $n, m \geq 1$. Se $f(X,Y)$ e $g(X,Y)$ não têm fator, não trivial, comum em $K[X,Y]$, então*

$$\#(V(f) \cap V(g)) \leq nm.$$

1.1.3 Multiplicidades

Vamos definir índice de interseção de uma curva plana afim com uma reta.

Sejam $f = 0$ e $\ell : Y = aX + b$ uma reta. Observe que

$$V(f \cap \ell) = \{(x,y) \in K^2; f(x, ax+b) = 0\}$$

Agora tomando $f_l(X) := f(X, aX + b)$ e usando propriedades elementares do anel principal $K[X]$, tem-se as seguintes possibilidades:

(1) $f_l(X) = 0$, nesse caso l é uma componente de f.

(2) $f_l(X) = c$, $c \in K^*$, quando $V(f \cap l) = \emptyset$.

(3) $f_l(X) \in K[X] \setminus K$, nesse caso:
$f_l = c\prod_{i=1}^{r}(X - x_i)^{m_i}, c \in K^*$ com $x_1, \ldots, x_r$ distintos em K.

Observação 1.1.2. *No caso particular das retas $\ell : X = c \ \in K$ evidentemente $f_l(Y) = f(c,Y)$ e todas as possibilidades acima são válidas trocando-se X por Y.*

Definição 1.1.5. *A multiplicidade ou índice de interseção de f e ℓ no ponto $P \in K^2$ é definido e denotado por:*

$$I(P, f \cap \ell) := \begin{cases} 0 & se \quad P \notin V(\ell \cap f), \\ \infty & se \quad P \in \ell, \ \ e\ \ell \ é\ uma\ componente\ de\ f, \\ m_i & se \quad P = (x_i, ax_i + b), \ \ 1 \leq i \leq r, \ \ onde\ \ f_l(X) = \prod_{j=1}^{r} (X - x_j)^{m_j}. \end{cases}$$

Se ℓ não é uma componente de f definimos

$$m_\infty = \partial f - \sum_{i=1}^{r} m_i := I(\infty, \ell \cap f)$$

Definição 1.1.6. *Seja $f = \sum_{0 \leq r+s \leq d} a_{rs} X^r Y^s \in K[X, Y]$, $d = \partial f$. Para cada i, $0 \leq i \leq d$ o polinômio $f_i = \sum_{r+s=i} a_{rs} X^r Y^s$ é chamado de componente homogênea de grau i de f e podemos escrever de maneira única : $f = f_0 + \ldots + f_d$ e $f_d \neq 0$, onde f_i são homogêneos de grau i de f.*

Como K é algebricamente fechado, $f_d(X, 1) = \prod_{a,b} (aX + b)$, e sendo f_d homogêneo, temos que $f_d(X, Y) = \prod_{a,b} (aX + bY)$.

Cada componente irredutível $aX + bY$ de f_d é dita direção assintótica de f.

O primeiro resultado a respeito de índice de interseção de curva com reta é o seguinte.

Proposição 1.1.2. *Seja $f = 0$ uma curva e $P \in V(f)$. Existe um inteiro $m = m_p(f) \geq 1$, tal que, para toda reta ℓ passando por P, temos $I(P, \ell \cap f) \geq m$, ocorrendo a desigualdade estrita para no máximo "m"retas e no mínimo uma.*

Demonstração: Suponhamos sem perda de generalidade, $P = (0, 0)$. Escrevamos $f = f_m + \ldots + f_d$ com f_i homogêneo de grau i para $m \leq i \leq d$ e $f_m \neq 0$.

Como $P \in V(f)$, temos $m \geq 1$. Mudando de coordenadas se necessário, podemos também supor que X não divide f_m. Daí, para a reta $X = 0$, $f(0, Y) = Y^m.[f_m(0, 1) + \ldots + f_d(0, 1).Y^{d-m}]$ e $f_m(0, 1) \neq 0$, donde conclui-se que $I(P, X \cap f) = m$.

Para as demais retas passando por $P = 0$, ponhamos $\ell_t = Y - tX$.

Temos então:

$$f(X, tX) = X^m.[f_m(1, t) + \ldots + f_d(1, t).X^{d-m}].$$

Segue que $I(P=0, \ell_t \cap f) \geq m$, ocorrendo a igualdade se, e somente se, $f_m(1,t) \neq 0$.

Como X não divide o polinômio homogêneo f_m, segue que $f_m(1,t)$ é um polinômio em t de grau m (≥ 1) e que portanto se anula para ao menos 1 e no máximo m valores distintos. $\square$

Definição 1.1.7. *O inteiro $m = m_p(f)$ descrito na proposição anterior é definido como sendo a multiplicidade do ponto P na curva f, ou simplesmente multiplicidade de f em P. Se $P \notin f$ convencionamos $m_p(f) = 0$.*

Dados uma curva f e um ponto $P = (x, y) \in V(f)$, considere $\tilde{f} = f(X + x, Y + y)$ e observe que um ponto $(a, b) \in V(f)$ se, e somente se, o ponto $(a - x, b - y) \in V(\tilde{f})$. Em particular, $0 = (0, 0) \in V(\tilde{f})$ e $m_p(f) = m_0(\tilde{f})$. Assim, podemos escrever $\tilde{f} = f(X + x, Y + y) = f_m(X, Y) +$ (termos de grau maior que m), com $f_m(X, Y)$ sendo a componente homogênea, não nula, de menor grau de $\tilde{f}$. Logo, se tomarmos $f_m = \prod_{e_i \geq 1} (a_i X - b_i Y)^{e_i}$, onde os fatores lineares $\tilde{\ell}_i = (a_i X - b_i Y)$ são retas distintas, podemos ver que $I(0, \tilde{\ell}_i \cap \tilde{f}_i) = I(P, \ell_i \cap f_i) > m_p(f)$, onde $\ell_i = a_i(X - x) - b_i(Y - y)$.

Definição 1.1.8. *A partir do que acabamos de descrever, definimos:*

(a) *As retas ℓ_i são denominadas retas tangentes de f em P.*

(b) *O expoente e_i é chamado de multiplicidade da reta tangente ℓ_i.*

(c) *O cone tangente de f em P é a união das retas tangentes, isto é, $CT_P f = \bigcup_t \ell_t$.*

Definição 1.1.9. *Seja $P \in f$, dizemos que P é*

(a) *liso (simples, não singular) ou que f é lisa em P, se $m_p(f) = 1$; singular, caso contrário.*

(b) *A curva f é lisa ou não singular se $m_p(f) = 1$ para cada $P \in f$.*

(c) *Se $m_p(f) = 2, 3, \ldots, m$, P é dito duplo, triplo, ..., m-uplo.*

(d) *Um ponto m-uplo $P \in f = 0$ é ordinário se f admitir m tangentes distintas em P.*

(e) *Uma cúspide é um ponto duplo com tangentes coincidentes.*

(f) *Um nó é um ponto duplo ordinário.*

O próximo resultado nos dá uma caracterização para pontos não singulares.

Proposição 1.1.3. *Seja $f = 0$ uma curva.*

(i) Um ponto $P \in f$ é liso se, e somente se, ao menos uma das derivadas parciais f_X, f_Y não se anula em P.

(ii) Se $P = (a, b) \in f$ é liso então a (única) tangente a f em P é dada por

$$f_X(P)(X - a) + f_Y(P)(Y - b) = 0.$$

Demonstração: Seja $P = (a, b) \in f$, pela fórmula de Taylor temos:

$$f(X + a, Y + b) = f(a, b) + f_X(a, b)X + f_Y(a, b)Y + g(X, Y),$$

onde os termos de g têm grau maior ou igual a 2.

(i) Como $P = (a, b) \in f$, temos que $f(a, b) = 0$. Daí,

$$P \text{ é liso} \Leftrightarrow m_p(f) = 1 \Leftrightarrow f_X(a, b) \neq 0 \text{ ou } f_Y(a, b) \neq 0.$$

(ii) Como $f(X, Y) = f_m(X, Y)$+(termos de grau maior que m) e P é liso, temos que $m_p(f) = 1$

Daí, $f_1(X, Y) = f_X(a, b)(X - a) + f_Y(a, b)(Y - b)$.

Então, por definição, a reta $l = f_X(P)(X - a) + f_Y(P)(Y - b)$ é a única reta tangente a f em P. □

1.1.4 Plano Projetivo

Suponha que queremos estudar todos os pontos de interseção de duas curvas; considere por exemplo a curva $Y^2 = X^2 + 1$ e a reta $Y = \alpha X$. Se $\alpha \neq \pm 1$, elas se cruzam em dois pontos. Quando $\alpha = \pm 1$, elas não se interceptam. Mas nós podemos ampliar o plano de tal maneira que tais curvas se interceptem "no infinito". Para isso, vamos definir plano projetivo.

Definição 1.1.10. *Vamos definir em K^3 a seguinte relação de equivalência:*

Dados $P, Q \in K^3 \setminus \{0\}$

$$P \sim Q \text{ se, e somente se, existe } \lambda \in K^* \text{ tal que } Q = \lambda P$$

Notação: Denotemos a classe de equivalência de $(x, y, z) \in K^3 \setminus \{0\}$ por $(x : y : z)$ e, definimos o plano projetivo por

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

Se $P = (x : y : z) \in \mathbb{P}^2_{\mathrm{K}}$ com $z \neq 0$ então, $P = P' = \left(\frac{x}{z} : \frac{y}{z} : 1\right)$.

O conjunto $U = \{(x : y : 1) \, ; \, x, y \in K\}$ é chamado espaço afim.

O conjunto $H_\infty = \{(x : y : 0) \, ; \, x, y \in K\}$ é chamado hiperplano infinito, e os pontos de H_∞ são chamados pontos no infinito.

Define-se analogamente os espaços afins e hiperplanos no infinito:

$U_1 = \{(1 : y : z) \, ; \, y, z \in K\}$, $H_{1\infty} = \{(0 : y : z) \, ; \, y, z \in K\}$.

$U_2 = \{(x : 1 : z) \, ; \, x, z \in K\}$, $H_{2\infty} = \{(x : 0 : z) \, ; \, x, z \in K\}$.

Observação 1.1.3. *Define-se* $\mathbb{P}^1_{\mathrm{K}}$ *por:* $\dfrac{K^2 \setminus \{0\}}{\sim}$.

Quando $K = \mathbb{R}$ *podemos visualizar* $\mathbb{P}^1_{\mathrm{K}}$ *como* $S^1 \approx \mathbb{R} \setminus \{\infty\}$.

1.1.5 Curvas Projetivas

Uma curva projetiva é uma curva em $\mathbb{P}^2_{\mathrm{K}}$.

Seja $F \in K[X, Y, Z]$ homogêneo de grau d. Tem-se:

$F(\lambda X, \lambda Y, \lambda Z) = \lambda^d F(X, Y, Z)$, com $\lambda \neq 0$.

Então, dado $(x, y, z) \in K^3$, $F(\lambda x, \lambda y, \lambda z) = 0$ se, e somente se, $F(x, y, z) = 0$.

Segue daí que se F é homogêneo, os zeros de F em $\mathbb{P}^2_{\mathrm{K}}$ independem do representante da classe de equivalência $\sim$. Assim, podemos definir:

Definição 1.1.11. *Dado um polinômio homogêneo de grau* $d \geq 1$*, o traço da curva* $F = 0$ *é o conjunto* $V(F) = \{P \in \mathbb{P}^2_{\mathrm{K}} \, ; \, \mathrm{F(P)} = 0$*, onde* $F \in K[X, Y, Z]\}$ *homogêneo.*

Exemplo 1.1.3. *Duas retas paralelas num plano afim, quando vistas em* $\mathbb{P}^2_{\mathrm{K}}$*, se interceptam no infinito.*

De fato, sejam $\ell = aX + bY + c$ e $\ell' = aX + bY + c'$ duas retas paralelas em U.

Homogeneizando ℓ e ℓ' obtemos $L = aX + bY + cZ$ e $L' = aX + bY + c'Z$.

Daí, $H_\infty(L) = \{(-b : a : 0)\} = H_\infty(L')$.

Logo, $(-b : a : 0) \in V(L) \cap V(L')$. □

Exemplo 1.1.4. *Seja $F = X^2Z - Y^2(Y - Z)$. Observe que se $U = V(F) \cap U$, temos:*

$V(F) \cap U = U(F) = \{(x, y) \in \mathbb{R}^2 ; y^3 = 0\}$ *e* $H_\infty(F) = \{(1 : 0 : 0)\}$

$U_1(F) = \{(y, z) \in \mathbb{R}^2 ; y^2(y - z) = 0\}$ *e* $H_{1\infty}(F) = \{(0 : 0 : 1), (0 : 1 : 1)\}$

$U_2(F) = \{(x, z) \in \mathbb{R}^2 ; x^2z = 0\}$ *e* $H_{2\infty}(F) = \{(1 : 0 : 0), (0 : 0 : 1)\}$. □

Definição 1.1.12. *Seja $F \in K[X, Y, Z]$ homogêneo, o polinômio $F_*(X, Y, Z) := F(X, Y, 1)$ é o que chamamos de desomogeneização de F.*

Por outro lado, dado $f \in K[X, Y]$ escreva $f(X, Y) = f_0(X, Y) + \ldots + f_d(X, Y)$ com f_i homogêneo de grau i. O polinômio $f^* = Z^d.f\left(\frac{X}{Z}, \frac{Y}{Z}\right) = f_0(X, Y)Z^d + \ldots + f_{d-1}(X, Y)Z + f_d(X, Y)$ é o que chamamos de a homogeneização de f.

Gostariamos de observar aqui que se $F \in K[X, Y, Z]$ é homogêneo e $V(F) \subseteq \mathbb{P}^2_K$ é o traço da curva projetiva então $V(F) \cap U = U(f) = V(F_*)$.

Por outro lado se $f \in K[X, Y]$ é não constante e $F = f^*$ então $V(F) \cap U = U(f)$, pois $(f^*)_* = f$ e $(F_*)^* = F$.

1.1.6 Interseção de reta e curvas projetivas

Aqui, nós vamos reproduzir para curvas projetivas, definições e resultados a respeito de interseção de retas e curvas que foram apresentados para curvas planas afins.

Assim, neste parágrafo, quando falarmos em reta e curvas estaremos nos referindo a retas e curvas projetivas. Assumiremos também neste parágrafo que K é um corpo algebricamente fechado.

Proposição 1.1.4. *Sejam L a reta $L = X = 0$ e $F = 0$ uma curva de grau d.*
Tome $P = (x : y : z) \in \mathbb{P}^2_K$. Então:

1. *$P = (x : y : z) \in L \cap F \iff x = 0$ e $F(0, y, z) = 0$*
2. *Se L divide F então, $V(L) \subset V(F)$*
3. *Se L não divide F então temos que:*

$$F(0,Y,Z) = \prod_{i=1}^{r}(z_iY - y_iZ)^{m_i}$$

com $V(F) \cap V(X) = \{P_1, \ldots, P_r\}$ *e* $P_i = (0 : y_i : z_i)$, $i = 1, \ldots, r$.

O expoente m_i *é chamado de* multiplicidade de interseção *de* X *e* F *em* P_i.

Demonstração: Os ítens 1 e 2 são claros. Vamos então provar o ítem 3.

Para cada i, tome a reta afim $\ell_i = z_iY - y_iZ$ e observe que dado $Q = (y, z) \in K^2 \setminus \{(0,0)\}$, o ponto $Q \in \ell_i$ se, e somente se, $F(0, y, z) = 0$.

De fato, se supomos $y \neq 0$, temos que $(y, z) \in \ell_i$ se, e somente se, $z_iy = y_iz$ e $y_i \neq 0$ (lembre que $(y_i, z_i) \in K^2 \setminus \{(0,0)\}$). E como $y^d.F(0, y_i, z_i) = F(0, yy_i, yz_i) = y_i^d.F(0, y, z)$, temos o que queríamos, isto é, ℓ_i divide $F(0, y, z)$.

Agora tome para cada i, $m_i = max\{n_i, \ell_i^{n_i} \setminus F(0, y, z)\}$.

Vamos ter que $F(0,Y,Z) = \left(\prod_{i-1}^{n} \ell_i^{m_i}\right).g(Y,Z)$, com $g(Y,Z)$ um polinômio homogêneo de grau r e para todo i, ℓ_i não divide g, mas se supomos que $r \geq 1$ e sendo K corpo algebricamente fechado, vamos ter que existe $(x, y) \in K^2 \setminus \{(0,0)\}$ tal que $g(y, z) = 0$, isto é, $F(0, y, z) = 0$.

Portanto $P = (0 : y : z) \in X \cap V(F)$, isto é, existe i tal que $P_i = P$ e ℓ_i divide $g(Y, Z)$, o que é uma contradição. □

Antes de definirmos índice de interseção de uma reta projetiva genérica com uma curva, será necessário introduzirmos o conceito de projetividade.

Para isso, apresentaremos a seguinte proposição elementar e natural.

Proposição 1.1.5. *Sejam* $T : K^3 \longrightarrow K^3$ *um isomorfismo* K*-linear e* A *a matriz* 3×3 *de* T^{-1} *na base canônica. Então:*

(a) *O isomorfismo* T *induz um* K*-automorfismo homogêneo do anel* $K[X, Y, Z]$ *dado por*

$$\begin{aligned} T_\circ : K[X,Y,Z] &\longrightarrow K[X,Y,Z] \\ F(X,Y,Z) &\longrightarrow F((A(X,Y,Z)^t)^t) \end{aligned}$$

(b) *O isomorfismo* T *induz uma bijeção natural de* $\mathbb{P}^2_K$ *em* $\mathbb{P}^2_K$*, chamada de projetividade e dada por*

$$T(x : y : z) = (a : b : c) \text{ onde } (a, b, c) = T(x, y, z)$$

Demonstração: As provas de (a) e (b) são bastante simples e omitiremos aqui. □

Definição 1.1.13. *Dizemos que duas curvas projetivas $F = 0$ e $G = 0$ são projetivamente equivalentes se existe uma projetividade T de $\mathbb{P}^2_{\mathrm{K}}$ em $\mathbb{P}^2_{\mathrm{K}}$ tal que $G = T_\circ F$.*

Em particular, elas têm o mesmo grau.

Proposição 1.1.6. *Sejam $L = aX + bY + cZ = 0$ uma reta e $F = 0$ uma curva de grau d. Se $V(L) \not\subset V(F)$, isto é, L não divide F então, $V(F) \cap V(L) = \{P_1, \ldots, P_r\}$, onde $P_i \neq P_j$ para $i \neq j$ e existem únicos inteiros $m_i \geq 1$ tais que, se T é uma projetividade tal que $T_\circ L = X$ então,*

$(T_\circ F)(0, Y, Z) = \prod_{i=1}^{r} (z_i Y - y_i Z)^{m_i}$, *onde* $T(P_i) = (0 : y_i : z_i)$ *para* $i = 1, \ldots, r$.

Em particular, $\sum m_i = d$.

Demonstração: Tome T um isomorfismo K-linear de K^3 em K^3 tal que $T_\circ L = X$ e considere o diagrama de homomorfismos:

$$\begin{array}{ccc} K[X,Y,Z] & \xrightarrow{T_\circ} & K[X,Y,Z] \\ q \downarrow & & \downarrow \overline{q} \\ \dfrac{K[X,Y,Z]}{< L >} & \xrightarrow{\overline{T_\circ}} & K[Y,Z] \end{array}$$

onde

$$\begin{cases} q : g \longmapsto \overline{g} = g + < L > \\ \overline{q} : g(X,Y,Z) \longmapsto g(0,Y,Z) \\ \overline{T_\circ} : \overline{g}(X,Y,Z) \longmapsto (T_\circ g)(0,Y,Z) \end{cases}$$

Como $T_\circ$ é isomorfismo então, $\overline{T_\circ}$ também é isomorfismo.

Segue que $\dfrac{K[X,Y,Z]}{< L >} \backsimeq K[Y,Z]$. Sendo $K[Y,Z]$ domínio fatorial então, $\overline{F} = P_1^{n_1} \ldots P_s^{n_s}$, com P_i irredutíveis e $n_i \geq 1$, para todo $F \in K[X,Y,Z]$.

Como $T_\circ \overline{F}(X,Y,Z) = (T_\circ F)(0,Y,Z) = \prod_{i=1}^{r} (z_i Y - y_i Z)^{m_i}$, onde $P_i = (0 : y_i : z_i)$ tal que $V(F) \cap V(X) = \{P_1, \ldots, P_r\}$, segue que $r = s$ e $n_i = m_i$, a menos da ordenação. □

Agora já temos condições de definir índice de interseção de uma reta com uma curva.

Definição 1.1.14. *A multiplicidade ou índice de interseção da reta $L = 0$ com uma curva $F = 0$ no ponto P é definido por:*

$$I(P, L \cap F) := \begin{cases} \infty \text{ se } P \in L \subset F \\ 0 \text{ se } P \notin L \cap F \\ m_i \text{ se } P = P_i \text{ nas condições da proposição anterior.} \end{cases}$$

Observação 1.1.4. *A proposição acima diz que sempre podemos supor que P esteja a uma distância finita, para o cálculo de $I(P, L \cap F)$, ou seja:*

$$I(P, L \cap F) = I(P, L_* \cap F_*).$$

Assim, calculamos a multiplicidade de interseção de curvas projetivas com retas da mesma forma que calculamos índice de interseção de retas com curvas afins.

Definição 1.1.15.

- *Se $P \notin V(F)$ então, $m_p(F) = 0$.*
- *$P \in V(F)$ é simples (não singular, liso) se $m_p(F) = 1$.*
- *$P \in V(F)$ é singular (múltiplo) se $m_p(F) \geq 2$.*

  Se $m_p(F) = 2, 3, \ldots$ diremos que P é um ponto duplo, triplo,... (respectivamente).

Exemplo 1.1.5. *Sejam a cúbica $Y = X^3$ e $F = Z^2Y - X^3$.*

No ponto $P = (0 : 1 : 0)$ temos que:

$F_* = Z^2 - X^3$ então, $m_p(F_*) = 2$.

Além disso, $F(P) = 0$ e $Z = 0$ logo, $X^3 = 0$.

Daí, segue que $I(P, F \cap Z) = 3$ □

A proposição seguinte, nos fornece um critério para decidir quando um ponto é singular.

Proposição 1.1.7. *Seja $F = 0$ uma curva de grau d e seja $P \in \mathbb{P}^2_{\mathrm{K}}$. Então:*

(a) $dF = XF_X + YF_Y + ZF_Z$. (Fórmula de Euler)

(b) P é um ponto singular de F se, e somente se, $F_X(P) = F_Y(P) = F_Z(P) = 0$.

(c) Se P é um ponto liso de F então, a reta tangente a F nesse ponto é:

$$F_X(P)X + F_Y(P)Y + F_Z(P)Z = 0.$$

Demonstração:

(a) Sendo ambos os membros da igualdade lineares como funções de F, é suficiente verificar a fórmula quando F é um monômio da forma $X^iY^jZ^k$, com $i + j + k = d$.

$F_X = iX^{i-1}Y^jZ^k$

$F_Y = jX^iY^{j-1}Z^k$

$F_Z = kX^iY^jZ^{k-1}$

Daí, $XF_X + YF_Y + ZF_Z = iX^iY^jZ^k + jX^iY^jZ^k + kX^iY^jZ^k = (i+j+k)(X^iY^jZ^k) = dF$.

Portanto, $dF = XF_X + YF_Y + ZF_Z$.

(b) Suponhamos $P = (a : b : 1)$. Então, pela Fórmula de Taylor para $f = F_*$ em P, temos:

$$f(x, y) = f(a, b) + f_X(a, b)(X - a) + f_Y(a, b)(Y - b) + \frac{1}{2}f_{XX}(a, b)(X - a)^2 + \dots.$$

Assim, $m_p(F) > 1 \Leftrightarrow f(a, b) = 0, f_X(a, b) = 0$ e $f_Y(a, b) = 0 \Leftrightarrow F_X(P) = 0$, $F_Y(P) = 0$ e $F(P) = 0$.

Pelo ítem (a) temos $dF(P) = aF_X(P) + bF_Y(P) + F_Z(P)$, daí segue que

$$m_p(F) > 1 \Leftrightarrow F_X(P) = F_Y(P) = F_Z(P) = 0 \text{ e } F(P) = 0.$$

(c) A reta tangente a F em $P = (a : b : 1)$ é dada pela equação:

$$f_X(a, b)(X - aZ) + f_Y(a, b)(Y - bZ) = 0$$

Então, $f_X(a, b)X + f_Y(a, b)Y - Z(af_X(a, b) + bf_Y(a, b)) = 0$. Logo, pelo ítem (a) temos $f_X(a, b)X + f_Y(a, b)Y + Z(f_Z(a, b)) = 0$.

Portanto, $F_X(P)X + F_Y(P)Y + F_Z(P)Z = 0$. □

Agora enunciaremos a versão geral do teorema de Bézout. Para maiores detalhes (incluindo sua demonstrção) ver [2]

Teorema 1.1.2 (Teorema de Bézout)**.** *Duas curvas projetivas F, G sem componentes irredutíveis em comum, tem $(\partial F)(\partial G)$ pontos em comum contados com multiplicidade.*

1.2 Curva Projetiva Plana

Seja K um corpo algebricamente fechado e $\mathcal{C} = V(F) \subseteq \mathbb{P}^2_{\mathrm{K}}$ uma curva projetiva. Se K_0 é subcorpo de K e existe $F_0 \in K_0[X, Y, Z]$ homogêneo tal que $\mathcal{C} = V(F_0)$, dizemos que $\mathcal{C}$ está definida sobre K_0 e denotamos $\mathcal{C}(K_0) = \mathcal{C} \cap \mathbb{P}^2_{\mathrm{K}_0}$.

1.3 Adição de pontos numa cúbica não singular

Neste parágrafo, seja K um corpo algebricamente fechado e K_0 um subcorpo de K. Mais ainda, vamos supor que $\mathcal{C} = V(F)$ é uma cúbica não singular definida sobre K_0 e que $\mathcal{C}(K_0)$ é não vazio, com $F \in K_0[X, Y, Z]$.

Dados dois pontos $P, Q \in \mathcal{C}(K_0)$, a reta L que passa por P e Q intercepta $\mathcal{C}$ em um terceiro ponto, que denotaremos $P * Q \in \mathbb{P}^2_{\mathrm{K}}$ (quando ocorrer $P = Q$, consideremos L como a reta tangente a $\mathcal{C}$ em P).

Proposição 1.3.1. *Se $P, Q \in \mathcal{C}(K_0)$ então, $(P * Q) \in \mathcal{C}(K_0)$.*

Demonstração: Seja $F \in K_0[X, Y, Z]$ e suponhamos que a reta que une P e Q é dada por

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

onde $a, b, c \in K_0$ e pelo menos um deles é não nulo.

Consideremos, sem perda de generalidade, $c \neq 0$. Obtemos então,

$$L : Z = \alpha X + \beta Y,$$

onde $\alpha, \beta \in K_0$.

Sejam $P = (x_1 : y_1 : z_1)$, $Q = (x_2 : y_2 : z_2)$ e $(P * Q) = (x_3 : y_3 : z_3)$, onde $x_i, y_i, z_i \in K_0$, para $i = 1, 2$.

Substituindo a equação da reta no polinômio F, obtemos

$$H(X,Y) = F(X,Y,\alpha X + \beta Y) = \lambda(a_1X - b_1Y)(a_2X - b_2Y)(a_3X - b_3Y),$$

onde $\lambda \in K_0^*$ e para cada $i = 1, 2, 3$, $a_1 \in K_0^*$ ou $b_i \in K_0^*$.

Assim, mudando os índices se for necessário, obtemos

$$\begin{array}{rcccl} P & = & (x_1 : y_1 : z_1) & = & (b_1 : a_1 : \alpha b_1 + \beta a_1) \\ Q & = & (x_2 : y_2 : z_2) & = & (b_2 : a_2 : \alpha b_2 + \beta a_2) \\ P * Q & = & (x_3 : y_3 : z_3) & = & (b_3 : a_3 : \alpha b_3 + \beta a_3) \end{array}$$

Logo, temos que existe $t \in K^*$, tal que $x_1 = b_1 t$, $y_1 = a_1 t$ e $z_1 = (\alpha b_1 + \beta a_1)t$.

Suponhamos que $b_1 \in K_0^*$, então como $t \in K_0^*$ temos que $b_1, a_1 \in K_0$.

Analogamente temos que $b_2, a_2 \in K_0$. Daí, segue que

$$H(X,Y) = P(X,Y)(a_3X - b_3Y),$$

onde $H, P \in K_0[X,Y]$.

Fazendo $Y = 1$, temos que o polinômio $H(X) = H(X,1) \in K_0[X]$ se fatora em $K_0[X]$.

Logo, $(a_3X - b_3) \in K_0[X]$.

Portanto, $(P * Q) \in \mathcal{C}(K_0)$. □

Definição 1.3.1. *Fixemos um ponto $\mathcal{O}$ de $\mathcal{C}(K_0)$. Dados dois pontos $P, Q \in \mathcal{C}(K_0)$, obtemos pela proposição anterior um terceiro ponto $(P * Q) \in \mathcal{C}(K_0)$.*

A reta que une $(P * Q)$ e $\mathcal{O}$, intercepta $\mathcal{C}$ num terceiro ponto $(P * Q) * \mathcal{O} \in \mathcal{C}(K_0)$.

Definimos a soma de dois pontos $P, Q \in \mathcal{C}(K_0)$, como sendo

$$P + Q = (P * Q) * \mathcal{O} \in \mathcal{C}(K_0).$$

Vamos agora discutir uma proposição que nos diz sobre a estrutura algébrica de $(\mathcal{C}(K_0), +)$. Mas antes disso, vamos enunciar um resultado que será usado na sua demonstração. Para maiores detalhes, ver [2]

Proposição 1.3.2. *Sejam C_1 e C_2 duas cúbicas que se interceptam em 9 pontos $P_1, P_2, \ldots, P_9$. Seja D uma cúbica que passa por $P_1, \ldots, P_8$. Então, D também passa por P_9.*

Proposição 1.3.3. *O conjunto $\mathcal{C}(K_0)$ com a adição definida acima forma um grupo comutativo.*

Demonstração:

- A adição é comutativa, pois a reta que passa por P e Q é a mesma reta que passa por Q e P.

- O ponto $\mathcal{O}$ é o elemento neutro da adição.

  De fato, dado qualquer $P \in \mathcal{C}(K_0)$, a reta L que passa por P e $\mathcal{O}$, intercepta $\mathcal{C}$ num terceiro ponto $P * \mathcal{O}$. Logo, a reta que passa por $\mathcal{O}$ e $P * \mathcal{O}$ é a mesma reta L.

  Portanto $P + \mathcal{O} = (P * \mathcal{O}) * \mathcal{O} = P$.

- A reta tangente a $\mathcal{C}$ no ponto $\mathcal{O}$, intercepta $\mathcal{C}$ em mais um ponto R.

  Temos que

  $$-P = P * R$$

  De fato, como a reta que passa por R e $\mathcal{O}$ é tangente a $\mathcal{C}$ em $\mathcal{O}$, segue que $R * \mathcal{O} = \mathcal{O}$. Portanto,

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

- Dados $P, Q, R \in \mathcal{C}(K_0)$, provaremos que

  $$(P + Q) + R = P + (Q + R).$$

  Para provarmos a igualdade acima, basta mostrarmos que $(P + Q) * R = P * (Q + R)$, pois daí segue que

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

  Dados dois pontos $a, b \in \mathcal{C}(K_0)$, denotamos por

  $$L = L(a, b, a * b)$$

  a reta que passa pelos pontos a, b e $a * b$.

  Definimos as retas

$$\begin{aligned} l &= l(P, Q, P * Q) \\ m &= m(P + Q, R, (P + Q) * R) \\ n &= n(Q * R, \mathcal{O}, Q + R) \\ s &= s(Q, R, Q * R) \\ t &= t(P, Q + R, P * (Q + R)) \\ u &= u(P * Q, \mathcal{O}, P + Q) \end{aligned}$$

Agora, usando essas retas definimos as cúbicas projetivas $G = l.m.n$ e $H = s.t.u$.

A curva $V(H)$ contém 8 pontos da interseção $V(F) \cap V(G)$, os quais são

$$\{\mathcal{O}, P, Q, R, P * Q, P + Q, Q * R, Q + R\}.$$

Pelo teorema de Cayley-Bacharach, as curvas $V(F), V(G)$ e $V(H)$ têm 9 pontos em comum. Portanto,

$$(P + Q) * R = P * (Q + R).$$

□

1.4 Curvas Elípticas

Definição 1.4.1. *Sejam K um corpo algebricamente fechado e K_0 um subcorpo de K. Diremos que uma cúbica plana afim, $C = V(f)$, está dada na forma de Weierstrass, se*

$$C : f(X, Y) = Y^2 - X^3 - aX^2 - bX - c = 0,$$

onde $a, b, c \in K_0$.

Observe que dada a cúbica f na forma de Weierstrass, homogeneizando f, podemos obter a curva plana projetiva $\mathcal{C}$, dada na forma

$$\mathcal{C} : F(X, Y, Z) = Y^2Z - X^3 - aX^2Z - bXZ^2 - cZ^3 = 0.$$

Observe que a curva $\mathcal{C}$ possui um único ponto no infinito, $\mathcal{O} = (0 : 1 : 0)$.

Além disso, o ponto $\mathcal{O}$ é não singular, pois

$$F_X(\mathcal{O}) = F_Y(\mathcal{O}) = 0 \text{ e } F_Z(\mathcal{O}) = 1.$$

Logo, a reta tangente a $\mathcal{C}$ em $\mathcal{O} = (0 : 1 : 0)$ está dada pela reta no infinito $L_0 : Z = 0$. Mais ainda, a multiplicidade de interseção de $V(F)$ com a reta $V(L_0)$ no ponto $\mathcal{O}$ é 3.

Seguindo as notações acima temos o lema:

Lema 1.4.1. *Se $V(L)$ é uma reta que intercepta $\mathcal{C}$ exatamente nos pontos P, Q e R, então*

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

Demonstração: Como a reta intercepta $\mathcal{C}$ em P, Q e R, e sendo $I(\mathcal{O}, F \cap L_0) = 3$, temos que $P * Q = R$, $R * \mathcal{O} = \mathcal{O}$ e $\mathcal{O} * \mathcal{O} = \mathcal{O}$.

Daí, segue que

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

□

Tome agora o polinômio $g(X) = X^3 + aX^2 + bX + c$.

Observamos que, existe um ponto singular em C se, e somente se, existe $p_0 = (x_0, y_0) \in C$, tal que

$$f_X(p_0) = -(3x_0^2 + 2ax_0 + b) = 0 \quad \text{e} \quad f_Y(p_0) = 2y_0 = 0$$

e isso ocorre se, e somente se, os polinômios $g(X)$ e $g'(X)$ admitem uma raiz em comum.

Assim, como $\mathcal{O} = (0 : 1 : 0)$ é o único ponto no infinito (não singular), temos que as seguintes condições são equivalentes

(a) C é não singular

(b) O polinômio $g(X)$ não tem raiz múltipla

(c) o discriminante de $g(X)$ $D = -4a^3c + a\ 2b^2 + 18abc - 4b^3 - 27c^2$ é diferente de zero.

Observação 1.4.1. *Podemos considerar $\mathcal{C}$ como a curva plana afim C com o ponto adicional no infinito $\mathcal{O} = (0 : 1 : 0)$, ou seja,*

$$\mathcal{C} = \{P \in A_K^2 \,;\, f(P) = 0\} \cup \{(0 : 1 : 0)\}.$$

Dessa forma,

$$\mathcal{C}(K_0) = \{P \in A_K^2 \,;\, f(P) = 0\} \cup \{(0 : 1 : 0)\}$$

define com a adição de pontos um grupo comutativo com elemento neutro $\mathcal{O} = (0 : 1 : 0)$, *sempre que* C *seja uma cúbica não singular.*

Definição 1.4.2. *Seja* C *uma cúbica plana dada na forma de Weierstrass*

$$C : Y^2 = X^3 + aX^2 + bX + c,$$

onde $a, b, c \in K$.

Diremos que C, *ou a curva projetiva* $\mathcal{C}$ *associada a* C, *é uma curva elíptica, se* C *é não singular. Mais ainda, diremos que uma curva elíptica está definida sobre* K_0, *se* $a, b, c \in K_0$.

Observação 1.4.2. *Da Observação 1.4.1 segue que, se* C *é uma curva elíptica sobre* K_0, *então* $C(K_0)$ *juntamente com a adição de pontos é um grupo comutativo.*

Na proposição seguinte, apresentaremos uma fórmula algébrica explícita para calcular a soma de dois pontos de uma curva elíptica. Para facilitar os cálculos, consideremos $K = \mathbb{C}$ e $K_0 = k$ uma extensão de $\mathbb{Q}$.

Proposição 1.4.1. *Seja* C *uma curva elíptica sobre* k *dada pela equação*

$$C : Y^2 = f(X) = X^3 + aX^2 + bX + c.$$

Sejam $P = (x, y)$ *e* $P_0 = (x_0, y_0) \in C(k)$. *Então*

(a) O oposto de P_0 *é* $-P_0 = (x_0, -y_0)$.

(b) Se $x = x_0$, *então* $P = P_0$ *ou* $P = -P_0$.

(c) Se $x \neq x_0$, *então* $P + P_0 = (\lambda^2 - a - x - x_0, -\lambda(\lambda^2 - a - x - x_0) - v)$, *onde*

$$\lambda = \frac{y - y_0}{x - x_0} \quad e \quad v = \frac{y_0 x - y x_0}{x - x_0}.$$

(d) Se $y \neq 0$, *então* $P + P = (\lambda^2 - a - 2x, -\lambda(\lambda^2 - a - 2x) - v)$, *onde*

$$\lambda = \frac{f'(x)}{2y} = \frac{3x^2 + 2ax + b}{2y} \quad e \quad v = \frac{-x^3 + bx + 2c}{2y}.$$

A reta $Y = \lambda X + v$ *é a reta tangente a* C *em* P.

Demonstração:

(a) Suponhamos $P_0 \neq \mathcal{O}$ (caso contrário a afirmação é clara).

Dado $P_0 = (x_0, y_0) \in C(k)$, observe que $-P_0$ é um ponto finito, pois caso contrário, $-P_0 = \mathcal{O}$, e isto implica que $P_0 = \mathcal{O}$, o que é falso.

Suponhamos que $-P_0 = (x_0', y_0')$.

Como a multiplicidade de intersecão da reta $Z = 0$ com C no ponto $\mathcal{O}$ é 3, segue que $-P_0 = P_0 * \mathcal{O}$.

A reta que passa por P_0 e $\mathcal{O}$ é dada por $X = x_0$.

Substituindo $X = x_0$ na equação da curva, obtemos

$$Y^2 - x_0^3 - ax_0^2 - bx_0 - c = 0.$$

As raízes desse polonômio são y_0 e y_0'. Portanto $y_0 + y_0' = 0$.

Logo, $y_0' = -y_0$ e $x_0' = x_0$.

(b) Suponhamos $x = x_0$.

A reta vertical $X = x_0$ intercepta C no ponto infinito e em dois pontos finitos.

- Se $y_0 \neq 0$, então segue de (a) que os pontos finitos são P_0 e $-P_0$, e portanto $P = P_0$ ou $P = -P_0$.
- Se $y_0 = 0$, então $y = 0$, pois caso contrário existiriam três pontos finitos distintos, P_0, P e $-P$, o que é falso. Portanto $P_0 = P$.

(c) Suponhamos $x \neq x_0$.

A reta que passa por P e P_0 tem equação $Y = \lambda X + v$, onde

$$\lambda = \frac{y - y_0}{x - x_0} \quad \text{e} \quad v = \frac{y_0 x - y x_0}{x - x_0}.$$

Substituindo $Y = \lambda X + v$ na equação da curva, obtemos o polinômio

$$X^3 + (-\lambda^2 + a)X^2 + (-2v\lambda + b)X + (-v^2 + c)$$

com x e x_0 raízes.

Assim, a outra raiz do polinômio é dada por

$$x' = -(-\lambda^2 + a) - x - x_0.$$

Portanto, a reta $Y = \lambda X + v$, intercepta C nos pontos P, P_0 e $P' = (x', \lambda x' + v)$.

Daí segue que $P + P_0 + P' = \mathcal{O}$.

Por (a), temos que

$$P = P_0 = -P' = (x', -\lambda x' - v) = (\lambda^2 - a - x - x_0, -\lambda(\lambda^2 - a - x - x_0) - v).$$

(d) Suponhamos que $y \neq 0$.

Se $P + P = \mathcal{O}$, então pelo item (a) temos que $y = -y$, donde segue que $y = 0$, o que é falso.

Logo, (P+P) é um ponto finito.

A inclinação da tangente a C no ponto P, pode ser achada derivando implicitamente, obtendo

$$\lambda = \frac{f'(x)}{2y} = \frac{3x^2 + 2ax + b}{2y}$$

Assim, a reta tangente a C em P, e dada por

$$Y = \lambda X + (y - \lambda X).$$

Sendo $P = (x, y)$ um ponto da curva, temos que

$$y^2 = x^3 + ax^2 + bx + c.$$

Logo,

$$v = y - \lambda x = \frac{2y^2 - x(3x^2 + 2ax + b)}{2y} = \frac{-x^3 + bx + 2c}{2y}.$$

Substituindo $Y = \lambda X + v$ na equação de C como foi feito em (c), e efetuando cálculos semelhantes, temos que

$$P + P = (\lambda^2 - a - 2x, -\lambda(\lambda^2 - a - 2x) - v).$$

□

Observação 1.4.3. *Seja C uma curva elíptica definida sobre K e m um inteiro. Podemos definir uma aplicação (aplicação multiplicação por m) $[m] : C(K) \longrightarrow C(K)$ dada por*

$$\begin{cases} [m](P) = \underbrace{P + P + \ldots + P}_{m\ vezes},\ se\ m > 0 \\ [m](P) = \mathcal{O}, se\ m = 0 \\ \\ [m](P) = \underbrace{(-P) + (-P) + \ldots + (-P)}_{-m\ vezes},\ se\ m < 0. \end{cases}$$

Essa aplicação, é a base para as operações em criptossistemas baseados em curvas elípticas.

1.5 Curvas Elípticas sobre corpos finitos

Vamos fazer algumas considerações sobre curvas elípticas definidas sobre corpos finitos, pois estas têm grandes aplicações na criptografia.

Nessa seção, vamos considerar K como o corpo finito $\mathbb{F}_q$ com $\#\mathbb{F}_q = q = p^r$

Seja C uma curva elíptica definida sobre $\mathbb{F}_q$.

Quando trabalhamos com corpos finitos, uma das primeiras coisas que é importante considerar, é o número de pontos da curva.

Uma primeira estimativa (cota superior) que temos, é q^2+1, a saber o ponto no infinito, mais q^2 possibilidades de pares $(x, y) \in \mathbb{F}_q$, já que temos q possibilidades para x e q possibilidades para y.

É fácil melhorar essa cota, basta observar que uma curva elíptica pode ter, no máximo, $2q + 1$, isto é, o ponto no infinito mais os $2q$ pares (x,y) com $x, y \in \mathbb{F}_q$. Pois para cada uma das q possibilidades para x, podem existir, no máximo 2 y satisfazendo a equação da curva.

Como apenas uma parte dos elementos de $\mathbb{F}_q^*$ têm raiz quadrada, é intuitivo esperar que o número de $\mathbb{F}_q$ pontos seja menor que $2q + 1$.

Para sermos mais precisos, seja χ o caracter quadrático de $\mathbb{F}_q$, ou seja, a aplicação que leva $x \in \mathbb{F}_q$ em ± 1, dependendo se x tem ou não raiz quadrada em $\mathbb{F}_q$ (convencionamos $\chi(0) = 0$).

Por exemplo, se $q = p$ é um número primo, então $\chi(x) = \left(\frac{x}{p}\right)$ é o Símbolo de Legendre.

Assim, o número de soluções para a equação $y^2 = u$ é igual a $1 + \chi(u)$. Logo o número de soluções (incluindo o ponto no infinito) é

$$1 + \sum_{x \in \mathbb{F}_q} (1 + \chi(x^3 + ax^2 + bx + c)) = q + 1 + \sum_{x \in \mathbb{F}_q} \chi(x^3 + ax + bx + c).$$

Nós podemos ainda ter uma melhor estimativa para o número de $\mathbb{F}_q$ pontos com o Teorema de Hasse que apenas enunciaremos aqui. Sua demonstração pode ser vista em [12]

Teorema 1.5.1. *Seja N o número de $\mathbb{F}_q$ pontos de uma curva elíptica definida sobre $\mathbb{F}_q$. Então:*

$$|N - (q + 1)| \leq 2\sqrt{q}.$$

Uma outra consideração importante a se fazer, é quanto à estrutura do grupo abeliano $(C(K), +)$. Esse grupo não é necessariamente cíclico, mas pode se mostrar que é sempre produto de dois grupos cíclicos.

Capítulo 2

Criptografia

Neste capítulo vamos introduzir o conceito de criptografia e apresentar alguns exemplos de sistemas criptográficos

2.0.1 Estimativa de tempo para os algorítmos

Antes de começarmos a falar sobre criptografia, vamos fazer uma pequena discussão a respeito de uma notação conveniente para estimar o tempo dos algorítmos.

Suponha que $f(n)$ e $g(n)$ são funções que tomam, para cada inteiro positivo n, valores positivos (mas não necessariamente inteiros). Dizemos que $f(n) = O(g(n))$ (ou simplesmente $f = O(g)$), isto é, o tempo de execução é da ordem $O(g(n))$, se existe uma constante C tal que $f(n)$ é sempre menor que $C.g(n)$. Por exemplo, $2n^2+3n-3 = O(n^2)$, a saber, $2n^2+3n-3 \leq 3n^2$.

Dizemos que um algorítmo é polinomial se o tempo de execussão é da ordem $O(n^k)$, onde k é constante. Além disso, qualquer algorítmo que não possa ser limitado por uma função polinomial é dito exponencial. Em geral, os algorítmos polinomiais são considerados eficientes, enquanto os exponenciais são ineficientes.

2.1 Noções básicas

Criptografia é o estudo de métodos para enviar mensagens de forma oculta, para que apenas receptores autorizados possam ler a mensagem.

A mensagem que queremos enviar é chamada de texto puro e chamamos de texto cifrado (ou codificado) a mensagem oculta.

Para escrevermos o texto puro e o texto cifrado, nós usamos um tipo de *alfabeto* consistindo

de um certo número N de letras. Aqui, o termo letra (ou caracter) é usado não apenas para as letras comuns A-Z do alfabeto tradicional, mas também refere-se a números, espaços em branco, pontuações, ou algum tipo de símbolo que quisermos. Vale observar que se não incluirmos o espaço em branco no *alfabeto*, as palavras são escritas todas juntas, tornando difícil a leitura da mensagem.

O processo de conversão do texto puro para o texto cifrado é chamado de cifragem e o processo reverso de decifragem.

O texto puro e o texto cifrado são quebrados em unidades de mensagem. As unidades de mensagem podem ser uma letra simples, um par de letras (dígrafo), uma tripla de letras (trígrafo) ou um bloco com N letras.

Uma função que toma uma unidade de mensagem de texto puro e leva em uma unidade de mensagem de texto cifrado é dita "transformação codificadora". Em outras palavras, é uma aplicação f que vai do conjunto α de todas as possíveis unidades de mensagem de texto puro no conjunto β de todas as possíveis unidades de mensagem de texto cifrado. Devemos sempre assumir que f é injetora, ou seja, dada uma unidade de mensagem de texto cifrado, existe uma, e apenas uma unidade de mensagem de texto puro que é levada nela. A "transformação"decodificadora"é a aplicação f^{-1} que recupera o texto puro do texto cifrado.

Observe a situação representada pelo diagrama seguinte:

$$\alpha \overset{f}{\longrightarrow} \beta \overset{f^{-1}}{\longrightarrow} \alpha$$

Qualquer configuração desse tipo é chamada de criptossistema.

Os primeiros passos para cifrar um criptossistema é "identificar"todas as possíveis unidades de mensagem de texto puro e cifrado com objetos matemáticos para que as funções possam ser construídas. Esses objetos normalmente são simplesmente números inteiros em alguma ordem. Por exemplo, se nossas unidades de mensagem de texto puro e cifrado são letras simples e tomamos como nosso *alfabeto* o alfabeto tradicional com 26 letras A-Z, nós podemos identificar as letras usando os inteiros $0, 1, 2, \ldots, 25$, os quais chamamos de números equivalentes. Assim, no lugar de A escrevemos 0, no lugar de S escrevemos 18, no lugar de X, 23 e assim por diante. Como um outro exemplo temos a seguinte situação, se nossas unidades de mensagem são dígrafos e tomamos o *alfabeto* com 27 letras constituindo do alfabeto tradicional A-Z, e do espaço em branco. Nós primeiro identificamos o espaço em branco com o número equivalente 26, e então identificamos os dígrafos cujas letras correspondem a $x, y \in \{0, 1, \ldots, 26\}$ com o

inteiro $27x + y \in \{0, 1, \ldots, 728\}$. Assim, nós vemos as letras individualmente como dígitos na base 27, e vemos os dígrafos como um 2-dígito inteiro nessa base.

Por exemplo, o dígrafo "NO"corresponde ao número $27.13 + 14 = 365$.

Analogamente, se nossas unidades de mensagens forem trígrafos, podemos identificá-las com os inteiros $729x + 27y + z \in \{0, 1, \ldots, 19682\}$

Em geral, podemos identificar blocos com K letras em um *alfabeto* com N letras com os inteiros entre 0 e $N^K - 1$, considerando cada bloco como um K-dígito inteiro na base N.

Podemos também identificar as unidades de mensagem usando outros objetos matemáticos como vetores ou pontos de alguma curva.

Vamos agora apresentar alguns exemplos. Começaremos com o caso onde tomamos como unidade de mensagem de texto puro e de texto cifrado letras simples em um *alfabeto* com N letras identificadas com os inteiros $0, 1, \ldots, N - 1$. Assim, a transformação codificadora é um reordenamento desses N inteiros.

Para facilitar a cifragem e a decifragem, é conveniente termos uma regra relativamente simples para reordenar os inteiros. Um forma de fazer isso, é ver o conjunto $\{0, 1, \ldots, N - 1\}$ como $\mathbb{Z}/N\mathbb{Z}$, e usar as operações de adição e multiplicação módulo N.

Exemplo 2.1.1. *Suponha que estamos usando o alfabeto com 26 letras constituindo das letras do alfabeto tradicional A-Z com números equivalentes 0-25. Seja* $x \in \{0, 1, 2, \ldots, 25\}$ *uma unidade de mensagem de texto puro. Defina a função* f*, que vai do conjunto* $\{0, 1, 2, \ldots, 25\}$ *nele mesmo, da seguinte forma:*

$$f(x) = \begin{cases} x + 3, & \text{se } x < 23 \\ x - 23, & \text{se } x \geq 23 \end{cases}$$

Em outras palavras, f *adiciona* 3 *módulo 26, ou seja,* $f(x) \equiv x + 3 \pmod{26}$.

Assim, com esse sistema, para cifrar a palavra "YES", o primeiro passo é converter para os números equivalentes 24 4 18, então agora adicionamos 3 módulo 26 e obtemos 1 7 21, que são os números equivalentes às letras "BHV". Logo, "BHV é a palavra "YES"cifrada.

Para decifrar uma mensagem, nós subtraímos 3 módulo 26. Por exemplo, o texto cifrado "ZKB"corresponde ao texto puro "WHY". De fato, "ZKB"tem 25 10 1 como números equivalentes. Subtraindo 3 módulo 26 obtemos 22 7 23 que correspondem a "WHY". Esse criptossistema foi aparentemente usado na Roma Antiga, por Julio César que ,supostamente, o teria

inventado. □

Podemos generalizar o exemplo 1 da seguinte forma. Suponha que estamos usando um *alfabeto* com N letras e números equivalentes $0, 1, \ldots, N-1$. Seja b um inteiro fixo. Definimos então a trasformação codificadora f da seguinte forma, $y = f(x) \equiv x + b \pmod{N}$. Chamamos essa transformação de transformação deslocamento. No caso onde $N = 26$ e $b = 3$ temos o criptossistema de Julio César. Para decifrar uma unidade de mensagem de texto cifrado $y \in \{0, 1, \ldots, N-1\}$, calculamos $x = f^{-1}(y) \equiv y - b \pmod{N}$.

Agora, suponha que nós não conhecemos as informações a respeito da cifragem e da decifragem, mas queremos ler a mensagem. Isso é chamado de quebra de código, e a ciência que estuda isso é a criptoanálise.

Para quebrar um criptossistema e conseguirmos ler a mensagem, são necessários dois tipos de informações. O primeiro é a natureza geral (a estrutura) do sistema. Por exemplo, suponha que sabemos que o criptossistema usa uma transformação deslocamento, letras simples e um *alfabeto* com 26 letras A-Z cujos números equivalentes são 0-25, respectivamente. O segundo tipo de informação necessária diz respeito a alguns parâmetros específicos do criptossistema. No nosso exemplo, o segundo tipo de informação que precisamos saber é a escolha do parâmetro b. Com estas informações podemos cifrar e decifrar as mensagens usando as fórmulas $y \equiv x + b \pmod{N}$ e $x \equiv y - b \pmod{N}$.

O parâmetro b é chamado de chave, ou mais precisamente, chave codificadora.

Nós vamos sempre assumir que a estrutura geral do sistema é conhecida.

Vamos apresentar agora um exemplo, onde o parâmetro b não é conhecido.

Exemplo 2.1.2. *Suponha que interceptamos a mensagem "FQOCUDEM", e que nós sabemos que foi cifrada usando uma transformação deslocamento, letras simples e um alfabeto com 26 letras como no exemplo anterior. Temos que descobrir o parâmetro b. Um caminho é através da análise de frequência, que se dá da seguinte forma. Suponha que interceptamos um longo cordão de texto cifrado, digamos com várias centenas de letras. Nós sabemos que "E"é a letra que ocorre com mais frequência na língua inglesa. Então, é razoável assumir que a letra que mais ocorre no texto cifrado é a codificação da letra "E".*

Suponha que "U"é o caracter que mais aparece no texto cifrado. Isso significa que a transformação deslocamento toma "E"=4 e leva em "U"=20, isto é, $20 \equiv 4 + b \pmod{26}$*, o que nos dá* $b = 16$*. Para decifrar a mensagem, então, basta subtrair 16 (trabalhando módulo 26) dos números equivalentes da mensagem "FQOCUDEM".*

$$\text{"}FQOCUDEM\text{"} = 5\ 16\ 14\ 2\ 20\ 3\ 4\ 12 \mapsto 15\ 0\ 24\ 12\ 4\ 13\ 14\ 22 = \text{"}PAYMENOW\text{"}$$

□

Nos caso da codificação usando uma transformação deslocamento, letras simples e um *alfabeto* com 26 letras, não é necessário ter um cordão longo da mensagem cifrada para encontrar a letra que ocorre mais freqentemente, pois existem apenas 26 possibilidades para b, e podemos simplesmente testar todas. Muito provavelmente, apenas uma resultará em uma mensagem que tenha sentido.

Esse tipo de criptossistema é bem simples e e também bem fácil de ser quebrado. Uma forma de melhorá-lo é usar um tipo mais geral de transformação em $\mathbb{Z}/N\mathbb{Z}$, chamada de aplicação afim e definida da seguinte forma: $y \equiv ax + b \pmod{N}$, onde a e b são inteiros fixos (juntos eles formam a chave codificadora). Por exemplo, trabalhando novamente com o *alfabeto* com 26 letras, se queremos cifrar a mensagem "PAYMENOW"usando a aplicação afim com chave codificadora $a = 7$, $b = 12$, obtemos:

$$\text{"}PAYMENOW\text{"} = 15\ 0\ 24\ 12\ 4\ 13\ 14\ 22 \mapsto 13\ 12\ 24\ 18\ 14\ 25\ 6\ 10 = \text{"}NMYSOZGK\text{"}$$

Para decifrar uma mensagem que foi cifrada usando uma aplicação afim $y \equiv ax+b \pmod{N}$, simplesmente colocamos x em função de y, obtendo $x \equiv a'y + b' \pmod{N}$, onde a' é o inverso de a módulo N e b' é igual a $-a^{-1}b$.

Note que isso vale apenas se $mdc(a, N) = 1$, caso contrário, não podemos colocar x em função de y. Se $mdc(a, N) > 1$, então é fácil ver que mais que uma letra do texto puro é levada na mesma letra do texto cifrado, e então não podemos recuperar unicamente a mensagem pura da mensagem cifrada. Aqui vale lembrar que sempre exigimos que a aplicação seja injetora.

Para resumir, um criptossistema afim em um *alfabeto* com N letras e com parâmetros $a \in (\mathbb{Z}/N\mathbb{Z})^*$ e $b \in \mathbb{Z}/N\mathbb{Z}$ consiste das seguintes regras:

$$y \equiv ax + b \pmod{N}, \qquad x \equiv a'y + b' \pmod{N}$$

onde

$$a' = a^{-1} \text{ em } (\mathbb{Z}/N\mathbb{Z})^*, \qquad b' = -a^{-1}b.$$

Como um caso especial do criptossistema afim, podemos tomar $a = 1$, obtendo as trasformações deslocamento. Um outro caso especial é quando $b = 0$: $y \equiv ax \pmod{N}$, $x \equiv a^{-1}y \pmod{N}$. O caso $b = 0$ é chamado de transformação linear.

Agora suponha que conhecemos uma mensagem interceptada que foi cifrada usando uma aplicação afim, letras simples em um *alfabeto* com N letras. Queremos determinar a chave codificadora a, b para poder ler a mensagem. Vamos ver um exemplo que ilustra essa situação.

Exemplo 2.1.3. *Continuaremos trabalhando com nosso alfabeto de 26 letras. Suponha que sabemos que a letra que ocorre com mais frequência na mensagem cifrada é "K", e a segunda mafreqentente é "D". É razoável assumir que são as codificações de "E"e "T", respectivamente, que são as duas letras que ocorrem com mais frequência na língua inglesa. Então, substituindo as letras por seus números equivalentes e substituindo x e y na fórmula decodificadora, obtemos*

$$\begin{aligned} 10a' + b' &\equiv 4 \pmod{26} \\ 3a' + b' &\equiv 19 \pmod{26} \end{aligned}$$

Nós temos duas congruências e duas incógnitas a' e b'. Uma forma de resolver esse sistema é subtrair a segunda linha da primeira para eliminar b'. Assim, nós obtemos $7a' \equiv 11 \pmod{26}$ e portanto, $a' \equiv 7^{-1}11 \equiv 9 \pmod{26}$. Finalmente, obtemos b' substituindo o valor de a' na primeira congruência resultando $b' \equiv 4 - 10a' \equiv 18 \pmod{26}$. Então, a mensagem pode ser decifrada usando a fórmula $x \equiv 9y + 18 \pmod{26}$. □

A Álgebra Linear nos diz que n equações são suficientes para encontrar n incógnitas, apenas quando essas equações são linearmente independentes (ou seja, o determinante é diferente de zero). No próximo exemplo, vamos ver um caso em que não é possível descobrir a chave codificadora usando analisando apenas as duas letras mais freqentes.

Exemplo 2.1.4. *Suponha que temos um cordão de texto cifrado e sabemos que ele foi codificado usando uma transformação afim com letras simples em um alfabeto com 28 letras consistindo das letras A-Z, um espaço em branco, e o ponto de interrogação ?, onde as letras A-Z têm números equivalentes 0-25, espaço em branco=26, ?=27. A análise de frequência revela que as duas letras mais comuns na mensagem cifrada são "B"e "?", nessa ordem. Então, como as duas letras mais comuns em textos na língua inglesa escritos com com esse alfabeto com 28 letras são " "(espaço em branco) e "E", nessa ordem, é razoável assumir que "B"é a codificação de " "e que "?"é a codificação de "E". Isso nos da duas congruências $a' + b' \equiv 26 \pmod{28}$ e $27a' + b \equiv 4 \pmod{28}$. Subtraindo as duas congruências, obtemos $2a' \equiv 22 \pmod{28}$, que implica $a' \equiv 11 \pmod{14}$. Isto significa que $a' \equiv 11$ ou $25 \pmod{28}$, e $b' \equiv 15$ ou 1*

(mod 28), *respectivamente. Essas possibilidades nos dão duas transformações decodificadoras afins* $11y + 15$ *e* $25y + 1$ *e ambas nos dão " "e "E"como texto puro correspondendo a "B"e "?", respectivamente. Nesse ponto, nós podemos testar as duas possibilidades e ver qual nos dá uma mensagem que tenha sentido. Ou, podemos continuar com a análise de frequência e ver qual é a terceira letra mais comum na língua inglesa e na mensagem cifrada. Suponha que "I"é a terceira letra mais freqente no texto cifrado. Usando o fato de que "T"é a terceira letra mais comum na língua inglesa (em nosso alfabeto com 28 letras), obtemos a terceira congruência:* $8a' + b' \equiv 19 \pmod{28}$. *Essa informação extra é suficiente para determinar qual das aplicaçõs afins é a correta. Nós encontramos que é* $11y + 15$. □

2.1.1 Transformações dígrafas

Vamos supor agora, que nossas unidades de mensagens de texto puro e cifrado são blocos com duas letras, chamados dígrafos. Se nós tivermos um texto puro com um número ímpar de letras então, para obtermos um número inteiro de dígrafos nós acrescentamos uma letra extra no fim do texto. Nós escolhemos uma letra que possivelmente não causará confusão, tal como um espaço em branco se nosso *alfabeto* contiver o espaço em branco, ou então alguma outra letra que escolhermos.

Cada dígrafo é então associado a um número equivalente. Um caminho simples para isso é associar o dígrafo ao número $xN + y$, onde x é o número equivalente à primeira letra do dígrafo e y o número equivalente ao segundo e N é o número de letras do nosso *alfabeto*. Isso nos dá uma correspondência injetora entre o conjunto de todos os dígrafos no *alfabeto* com N letras e o conjunto de todos os inteiros não negativos menores que N^2. Nós descrevemos essa identificação dos dígrafos para o caso especial quando $N = 27$ na seção anterior

O próximo passo é construir uma tranformação codificadora, isto é, uma reordenação dos inteiros $\{0, 1, \ldots, N^2 - 1\}$. As transformações mais simples são as afins, onde vemos o conjunto dos inteiros como $\mathbb{Z}/N\mathbb{Z}$ e definimos a codificação de uma unidade de mensagem de texto puro x como um inteiro y não negativo menor que N^2 satisfazendo a congruência $y \equiv ax+b \pmod{N^2}$. Aqui, como antes, "a"não deve ter fator comum com N (o que implica que não tem fator comum com N^2), a fim de que tenhamos uma transformação inversa conhecida como transformação decodificadora e definida pela congruência $x \equiv a'y + b' \pmod{N^2}$, onde $a' \equiv a^{-1} \pmod{N^2}$ e $b' \equiv -a^{-1}b \pmod{N^2}$. Em seguida, nós transformamos y em um bloco de texto cifrado com duas letras escrevendo-o na forma $y = x'N + y'$ e então, para finalizar, vemos quais as letras que correspondem aos números equivalentes x' e y'.

Exemplo 2.1.5. *Nós sabemos que nosso adversários estão usando um criptossistema com um alfabeto com 27 letras, consistindo das letras A-Z, que têm números equivalentes* $0-25$*, respectivamente, e do espaço em branco com número equivalente 26. Cada dígrafo corresponde então a um inteiro entre* 0 *e* $728 = 27^2 - 1$*. Suponha que o estudo de um grande cordão de texto cifrado revelou que os dígrafos que ocorrem com mais frequência são, em ordem, "ZA", "IA"e "IW". Suponha que os dígrafos mais comuns na língua inglesa são "E..."(isto é, E e um espaço em branco), "S..."e "...T"Nós sabemos que o criptossistema usado foi um afim com transformação codificadora módulo 729. Encontre a chave decodificadora, leia a mensagem "NDXBHO"e encontre a chave codificadora.*

Solução: Nós sabemos que o texto puro é cifrado usando a regra $y \equiv ax + b \pmod{729}$, e o texto cifrado pode ser decodificado com a regra $x \equiv a'y + b' \pmod{729}$, onde a e b formam a chave codificadora e, a' e b' formam a chave decodificadora. Nós primeiro queremos encontrar a' e b'. Sabemos como três dígrafos são decodificados, e, depois de substituir os dígrafos por seus números equivalentes, nós temos as seguintes congruências:

$$\begin{aligned} 675a' + b' &\equiv 134 \pmod{729}, \\ 216a' + b' &\equiv 512 \pmod{729}, \\ 238a' + b' &\equiv 721 \pmod{729}. \end{aligned}$$

Subtraindo a terceira congruência da primeira obtemos $437a' \equiv 142 \pmod{729}$. Para resolver isso, temos que encontrar o inverso de 437 módulo 729. Para isso vamos usar o algoritmo de Euclides:

$$\begin{aligned} 729 &= 437 + 292 \\ 437 &= 292 + 145 \\ 292 &= 2.145 + 2 \\ 145 &= 72.2 + 1 \end{aligned}$$

e então,

$$\begin{aligned} 1 &= 145 - 72.2 \\ &= 145 - 72(292 - 2.145) \\ &= 145.145 - 72.292 \\ &= 145(437 - 292) - 72.292 \\ &= 145.437 - 217.292 \\ &= 145.437 - 217(729 - 437) \\ &= 362.437 \pmod{729} \end{aligned}$$

Então, $a' \equiv 362.142 \equiv 374 \pmod{729}$ e portanto $b' \equiv 134 - 675.374 \equiv 647 \pmod{729}$. Agora, aplicando a transformação decodificadora nos dígrafos "ND", "XB"e "HO"da mensagem que queremos ler (eles correspondem aos inteiros 354, 622 e 203, respectivamente), nós obtemos os inteiros 365, 724 e 24. Escrevendo 365=13.27 + 14, 724=26.27 + 22 e 24=0.27 + 24 e substituindo os números pelas letras equivalentes nós temos o texto puro "NO WAY". Finalmente, para encontrar a chave codificadora, basta calcular $a \equiv a'^{-1} \equiv 374^{-1} \equiv 614 \pmod{729}$ (novamente usando o algoritmo de euclides) e $b \equiv -a'^{-1}b' \equiv 614.647 \equiv 47 \pmod{729}$. □

2.1.2 Matrizes codificadoras

Suponha que estamos trabalhando com um *alfabeto* com N letras e queremos enviar mensagens usando dígrafos como unidades de mensagem. Ao invés de considerarmos cada dígrafo como um inteiro considerado módulo N^2, ou seja, como um elemento de $\mathbb{Z}/N^2\mathbb{Z}$, como vimos anteriormente, podemos considerar cada dígrafo correspondendo a um vetor, isto é, a um par de inteiros $\binom{x}{y}$, onde x e y são considerados módulo N. Para exemplificar, se usarmos o *alfabeto* com 26 letras A-Z, com números equivalentes 0-25, respectivamente, então o dígrafo "NO"corresponde ao vetor $\binom{13}{14}$.

Nós olhamos cada dígrafo P como um ponto de $N \times N$. Isto é, temos um plano xy, exceto pelo fato de que os eixos não são cópias da reta real, e sim cópias de $\mathbb{Z}/N\mathbb{Z}$. Assim como o plano real xy normalmente é denotado por $\mathbb{R}^2$, $N \times N$ é denotado por $(\mathbb{Z}/N\mathbb{Z})^2$.

Uma vez que visualizamos dígrafos como vetores, também interpretamos uma transformação codificadora como um rearranjo dos pontos de $N \times N$. Mais precisamente, uma aplicação codificadora é uma função injetora que vai do conjunto $(\mathbb{Z}/N\mathbb{Z})^2$ nele mesmo.

Por muito tempo, um dos métodos mais populares de era o seguinte: para algum k fixado, considere blocos de k letras como vetores em $(\mathbb{Z}/N\mathbb{Z})^k$. Escolha algum vetor fixo $b \in (\mathbb{Z}/N\mathbb{Z})^k$ e codifique usando o vetor translação $C = P+b$, onde as unidades de mensagens de texto puro e cifrado são k-úplas de inteiros módulo N. Mas, como é fácil perceber, esse criptossistema pode ser quebrado com facilidade, pois se conhecemos (ou podemos supor) N e k, basta quebrarmos o texto cifrado em blocos com k letras e executar a análise de frequência nas primeiras letras de cada bloco para determinar a primeira componente de b. Então, depois fazemos a mesma coisa com as segundas letras de cada bloco, e assim por diante.

Observação 2.1.1. *Temos da Álgebra linear que dada a matriz* $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, *com* a, b, c *e* d *são números reais, a matriz é inversível se, e somente se, o determinante* $D = ad - bc$ *é diferente*

de zero. Quando trabalhamos em um anel arbitrário R, temos uma situação análoga. De fato, suponha que

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(R)$$

e $D = det(A) = ad - bc \in R^*$. Seja D^{-1} o inverso multiplicativo de D em R. Então:

$$\begin{pmatrix} D^{-1}d & -D^{-1}b \\ -D^{-1}c & D^{-1}a \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} D^{-1}(da - bc) & 0 \\ 0 & D^{-1}(-cb + ad) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Se multiplicarmos na ordem inversa, vamos obter o mesmo resultado.

Além dessa, temos outras condições suficientes para que uma matriz seja inversível. Isso pode ser visto no seguinte resultado:

Proposição 2.1.1. *Sejam $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbb{Z}/N\mathbb{Z})$ e $D = ad - bc$. As seguintes condições são equivalentes:*

(a) *$mdc(D, N) = 1$;*

(b) *A tem uma matriz inversa;*

(c) *se x e y não são ambos zero em $\mathbb{Z}/N\mathbb{Z}$, então $A\binom{x}{y} \neq \binom{0}{0}$;*

(d) *A nos dá uma correspondência injetora do conjunto $(\mathbb{Z}/N\mathbb{Z})^2$ com ele mesmo.*

Demonstração: *Já provamos que $(a) \Longrightarrow (b)$. Agora vamos provar que*

$$(b) \Longrightarrow (d) \Longrightarrow (c) \Longrightarrow (a)$$

Suponha que (b) é verdadeira. Então (d) também vale, pois A nos dá uma aplicação que leva $\begin{pmatrix} x \\ y \end{pmatrix}$ em $\begin{pmatrix} x' \\ y' \end{pmatrix} = A\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$ e A^{-1} nos dá a aplicação inversa que leva $\begin{pmatrix} x' \\ y' \end{pmatrix}$ em $\begin{pmatrix} x \\ y \end{pmatrix}$.

Agora, se vale (d), temos que $\begin{pmatrix} x \\ y \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ implica $A\begin{pmatrix} x \\ y \end{pmatrix} \neq A\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. Logo (c) também é verdadeira.

Agora suponha que (a) é falsa. Suponha também que $m = mdc(D, N) > 1$ e que $m' = \dfrac{N}{m}$. Temos três casos possíveis:

- *Se todas as quatro entradas de A são divisíveis por m, tome $\binom{x}{y} = \binom{m'}{m'}$ e temos uma contradição com (c).*

- *Se a e b não são ambos divisíveis por m, tome $\binom{x}{y} = \binom{-bm'}{am'}$. Assim,*

$$A\binom{x}{y} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\binom{-bm'}{am'} = \begin{pmatrix} -abm' + bam' \\ -cbm' + dam' \end{pmatrix} = \binom{0}{Dm'} = \binom{0}{0},$$

  pois como m divide D, segue que $N = mm'$ divide Dm'.

- *Se c e d não são ambos divisíveis por m, tome $\binom{x}{y} = \binom{dm'}{-cm'}$ e proceda como no caso anterior.*

Logo, em todos os casos temos uma contradição, o que implica que $(c) \Longrightarrow (a)$. □

Assim, pela observação 2.1.1, nós podemos obter transformações codificadoras dos nossos vetores usando matrizes $A \in M_2(\mathbb{Z}/N\mathbb{Z})$ cujo determinante não tenha fator comum com N.

A saber, cada unidade de texto puro $P = \binom{x}{y}$ é transformado em um texto cifrado $C = \binom{x'}{y'}$ pela seguinte regra

$$C = AP, \qquad \text{isto é} \qquad \binom{x'}{y'} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\binom{x}{y}.$$

Para decodificar uma mensagem, basta aplicar a matriz inversa:

$$P = A^{-1}AP = A^{-1}C, \qquad \text{isto é} \qquad \binom{x}{y} = \begin{pmatrix} D^{-1}d & -D^{-1}b \\ -D^{-1}c & D^{-1}a \end{pmatrix}\binom{x'}{y'}.$$

Um modo mais geral para codificar um vetor dígrafo $P = \binom{x}{y}$ é aplicar uma matriz 2×2 $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbb{Z}/N\mathbb{Z})$ e então adicionar um vetor constante $B = \binom{e}{f}$:

$$C = AP + B,$$

isto é,

$$\binom{x'}{y'} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\binom{x}{y} + \binom{e}{f} = \begin{pmatrix} ax + by + e \\ cx + dy + f \end{pmatrix}.$$

Essa aplicação é chamada de afim, e é análoga a função codificadora $y = ax + b$ que já foi descrita, onde nós usamos letras simples como unidade de mensagem.

Observação 2.1.2. *Aqui, quando usamos o sinal "="significa que as entradas correspondentes são congruentes módulo N.*

A transformação inversa, que expressa P em função de C pode ser obtida, subtraindo B de ambos os lados e depois aplicando A^{-1} também em ambos os lados.

$$P = A^{-1}C - A^{-1}B$$

2.2 Criptografia com chave pública

Lembremos que um criptossistema consiste de uma transformação codificadora injetora f que vai do conjunto α de todas as possíveis unidades de mensagem de texto puro no conjunto β de todas as possíveis unidades de mensagem de texto cifrado. Atualmente, o termo criptossistema é também usado para se referir a uma família de transformações codificadoras, cada uma correspondendo a uma escolha de parâmetros (os conjuntos α e β, assim como a aplicação f, dependem dos valores dos parâmetros). Por exemplo, para um *alfabeto* fixo com N letras (com números equivalentes também fixos), nós podemos considerar o criptossistema afim (ou "família de criptossistemas"), onde para cada $a \in (\mathbb{Z}/N\mathbb{Z})^*$ e $b \in \mathbb{Z}/N\mathbb{Z}$ temos a aplicação que vai de $\alpha = \mathbb{Z}/N\mathbb{Z}$ em $\beta = \mathbb{Z}/N\mathbb{Z}$ definida por $y \equiv ax+b \pmod{N}$. Nesse exemplo, os conjuntos α e β são fixos pois N é fixo, mas a transformação codificadora f depende da escolha dos parâmetros a e b. A transformação codificadora pode então ser descrita por um algoritmo, que é o mesmo para toda a família e pelos valores dos parâmetros. Os valores dos parâmetros são chamados de chave codificadora K_E. Em nosso exemplo, K_E é o par (a, b). Na prática, nós sempre supomos que o algoritmo é publicamente conhecido, ou seja, o processo geral usado para codificar não é mantido em segredo. No entanto, as chaves podem, facilmente, ser trocadas periodicamente.

Nós também precisamos de um algoritmo e de uma chave para poder decodificar, isso é, calcular f^{-1}. Essa chave é chamada de chave decodificadora K_D. Em nosso exemplo da família de criptossistemas afins, para decodificar também usamos uma aplicação, a saber $x = a^{-1}y - a^{-1}b \pmod{N}$, e então, as transformações decodificadoras usam o mesmo algoritmo que as transformações codificadoras, exceto pela chave, que é diferente, a saber, o par $(a^{-1}, -a^{-1}b)$. É importante observar, que em alguns criptossistemas, o algoritmo decodificador, assim como a chave, é diferente do algoritmo codificador. Nós sempre vamos supor que os algoritmos codificador e decodificador são publicamente conhecidos e que as chaves K_E e K_D podem ser escondidas.

Suponha agora, que alguém quer se comunicar secretamente usando o criptossistema afim $y \equiv ax+b$. Nós já observamos que não é difícil quebrar esse sistema se estiver sendo usado letras simples como unidades de mensagem em um *alfabeto* com N letras. É um pouco mais difícil de quebrar um sistema que usa dígrafos. Quanto maior o bloco de letras que for usado como unidade de mensagem, mais seguro é o sistema. Mas em todos os exemplos que apresentamos até agora, não é necessário especificar a chave decodificadora, uma vez que a chave codificadora é conhecida. Mesmo que nós estivermos trabalhando com números grandes (tais como N^k, com k bem grande), é possível determinar a chave decodificadora a partir da chave codificadora. Por exemplo, no caso da tranformação codificadora afim de $\mathbb{Z}/N^k\mathbb{Z}$, uma vez que nós conhecemos a chave codificadora $K_E = (a, b)$ nós podemos calcular a chave decodificadora $K_D = (a^{-1} \pmod{N^k}, -a^{-1}b \pmod{N^k})$ através do algoritmo de Euclides. Assim, com um criptossistema tradicional, qualquer um que saiba o suficiente para decifrar uma mensagem pode, com pouco esforço extra, determinar as chaves codificadora e decodificadora.

Há um outro tipo de criptossistema que é mais seguro no sentido de manter oculta a chave decodificadora. É o criptossistema com chave pública. A razão para o nome chave pública é que a informação necessária para enviar mensagens secretas - a chave codificadora K_E - pode ser uma informação pública (conhecida por todos) sem permitir com isso, que qualquer pessoa possa ler a mensagem.

Por definição, um criptossistema com chave pública tem a propriedade que alguém que sabe apenas como codificar não pode usar a chave codificadora para encontrar a chave decodificadora sem um cálculo extremamente longo. Em outras palavras, a função codificadora $f : \alpha \longrightarrow \beta$ é fácil de ser calculada uma vez que a chave codificadora K_E é conhecida, mas é muito difícil calcular a função inversa $f^{-1} : \beta \longrightarrow \alpha$. Isto significa, do ponto de vista computacional, que a função f não é invertível (sem alguma informação extra - a chave decodificadora K_D). Tal função é chamada de função "trapdoor".

Há um conceito parecido com a função trapdoor, que é a "função com sentido único". Esse tipo de função, é fácil de ser calculada, mas é muito difícil calcular a inversa f^{-1} e não há nenhuma informação adicional que a torne fácil de ser calculada. A noção função "trapdoor", aparentemente, apareceu pela primeira vez em 1978 junto com a invenção do criptossistema de chave pública RSA. Já, a noção de função com sentido único é mais velha. O que parece ter sido o primeiro uso de de funções com sentido único em criptografia foi descrito no livro de Wilkes que foi publicado em 1968.

Observe que em um sistema com chave pública é possível que duas pessoas se comuniquem

secretamente sem terem tido nenhum contato prévio, sem trocar nenhuma informação preliminar. Toda as informações necessárias para emitir uma mensagem cifrada estão disponíveis publicamente.

2.2.1 Autenticação

Frequentemente, uma das partes mais importantes da mensagem é a assinatura. Se a mensagem for particularmente importante ou em uma comunicação eletrônica, onde não é possível ter uma assinatura física, pode ser necessário usar métodos adicionais para autenticar a comunicação. Por exemplo, quando uma pessoa quer movimentar o dinheiro de sua conta telefone, é pedido frequentemente para dar alguma informação pessoal (por exemplo, o nome da mãe), que a pessoa e o banco sabem (dados que foram informados quando a conta foi aberta), mas que um impostor provavelmente não saberia.

Na criptografia com chave pública, há uma maneira especialmente fácil de identificar-se de modo que ninguém possa fingir ser você. Sejam Alice (A) e Bernardo (B) dois usuários do sistema. Seja f_A a transformação codificadora que qualquer usuário deve usar para enviar mensagens a Alice, e seja f_B o mesmo para Bernardo. Para simplificar, vamos assumir que o conjunto de todas as possíveis unidades de mensagem pura (α) e o conjunto de todas as possíveis unidades de mensagem cifrada (β) sejam iguais, e os mesmos para todos os usuários do sistema. Seja P a assinatura de Alice (talvez incluindo um número de identificação, ou qualquer informação que garanta que a mensagem é mesmo de Alice). Não é suficiente que Alice envie para Bernardo a mensagem cifrada $f_B(P)$, já que todos sabem como fazer isso e assim não haveria nenhuma maneira de saber que a assinatura não foi forjada. Então, no começo (ou no fim) da mensagem, Alice transmite $f_B f_A^{-1}(P)$. Assim, quando Bernardo decodificar toda a mensagem, incluindo essa parte, aplicando f_B^{-1}, ele verá que tudo foi transformado em unidade de mensagem pura, exceto uma pequena parte, que é $f_A^{-1}(P)$. Como Bernardo sabe que a mensagem é, supostamente, de Alice, ele aplica f_A (que ele conhece, pois a chave codificadora de Alice é pública), e obtém P. Como ninguém além de Alice poderia ter aplicado a função f_A^{-1}, que é invertida aplicando f_A, ele finalmente tem certeza que a mensagem veio de Alice.

2.3 RSA

Procurando uma função trapdoor f para usar em um criptossistema com chave pública, queremos usar uma idéia que seja razoavelmente simples e de fácil execução. Em outras palavras, queremos ter evidências fortes de que a decodificação não pode ser realizada sem o conheci-

mento da chave decodificadora. Por essa razão, é natural olhar um problema antigo da teoria dos números: o problema de encontrar a fatoração completa de um inteiro composto grande cujos fatores primos não são conhecidos anteriormente. O sucesso do chamado criptossistema RSA (RSA vem dos últimos nomes dos inventores Rivest, Shamir e Adleman), que é um dos mais velhos e mais populares criptossistemas com chave pública, é baseado na grande dificuldade de fatorar.

Vamos descrever como o RSA funciona. Primeiramente, cada usuário escolhe dois números primos extremamente grandes p e q (digamos, aproximadamente, 100 dígitos) e toma $n = pq$. Sabendo a fatoração de n, é fácil calcular $\varphi(n) = (p-1)(q-1) = n+1-p-q$. Em seguida, o usuário escolhe, aleatoriamente, um inteiro e entre 1 e $\varphi(n)$ que é primo com $\varphi(n)$.

Observação 2.3.1. *Quando dizemos "número aleatório", queremos dizer que o número foi escolhido com a ajuda de um gerador de números aleatórios, isto é, um programa de computador que gera uma sequencia de dígitos de maneira que ninguém poderia saber qual seria o próximo dígito. No criptossistema RSA nós precisamos de um gerador de números aleatórios não apenas para escolher e, mas também para escolher os primos grandes p e q (de modo que ninguém poderia adivinhar nossas escolhas olhando para tabelas de tipos especiais de números primos, como por exemplo os primos de Mersenne). O que um número primo gerado aleatoriamente significa? Bem, primeiro gere um inteiro grande aleatório m. Se m é par, substitua por $m+1$. Então, aplique um teste de primalidade apropriado para ver se o número ímpar m é ou não primo. Se m não é primo, tente $m+2$, depois $m+4$, e assim por diante, até encontrar o primeiro número primo maior ou igual a m.*

Similarmente, o número aleatório "e"que é primo com $\varphi(n)$ pode ser escolhido gerando-se primeiramente um inteiro aleatório depois, ir sucessivamente incrementando-o até que se encontre um e com a propriedade $mdc(e, \varphi(n)) = 1$.

Assim, cada usuário A escolhe dois primos p_A e q_A e um número aleatório e_A que não tenha fator comum com $(p_A-1)(q_A-1)$. Em seguida, A calcula $n_A = p_A q_A$, $\varphi(n_A) = n_A+1-p_A-q_A$, e também o inverso multiplicativo de e_A módulo $\varphi(n_A)$: $d_A = e_A^{-1} \pmod{\varphi(n_A)}$. Então, ele torna público a chave codificadora $K_{E,A} = (n_A, e_A)$ e mantém secreta a chave decodificadora $K_{D,A} = (n_A, d_A)$. A transformação codificadora é uma aplicação que vai do conjunto $\mathbb{Z}/n_A\mathbb{Z}$ nele mesmo e é dada por $f(x) \equiv x^{e_A} \pmod{n_A}$. A transformação decodificadora vai do conjunto $\mathbb{Z}/n_A\mathbb{Z}$ nele mesmo e é dada por $f^{-1}(y) \equiv y^{d_A} \pmod{n_A}$. Não é difícil ver que uma função é a inversa da outra.

Como podemos ver no parágrafo anterior, estamos trabalhando com os conjuntos das unida-

des de mensagens de texto puro e texto cifrado iguais $\alpha = \beta$. Na prática, nós queremos escolher α e β uniformemente durante todo o sistema. Por exemplo, suponha que estamos trabalhando em um *alfabeto* com N letras. Então, sejam $k < l$ inteiros positivos escolhidos apropriadamente, tais que, por exemplo, N^k e N^l tenham aproximadamente 200 dígitos. Tomamos então, blocos com k letras como nossas unidades de mensagem de texto puro, que consideramos como k-dígitos na base N, isto é, nós atribuímos a eles números equivalentes entre 0 e N^k. Similarmente, tomamos como nossas unidades de mensagem de texto cifrado blocos com l letras. Então, cada usuário deve escolher seus primos grandes p_A e q_A tais que $n_A = p_A q_A$ satisfaça $N^k < n_A < N^l$. Assim, qualquer unidade de mensagem de texto puro, isto é, inteiros menores que N^k, correspondem a um elemento de $\mathbb{Z}/n_A\mathbb{Z}$; e como $n_A < N^l$, a imagem $f(P) \in \mathbb{Z}/n_A\mathbb{Z}$ pode ser escrita, de maneira única, como um bloco com l letras.

Vamos agora ver um exemplo de como o criptossistema RSA funciona.

Exemplo 2.3.1. *Como é apenas para vermos como o criptossistema funciona, vamos trabalhar com inteiros relativamente pequenos.*

Escolha $N = 26, k = 3$ e $l = 4$. Isto é, o texto puro consiste em trígrafos e o texto cifrado consiste em blocos com 4 letras no alfabeto usual com 26 letras. Para enviar a mensagem "YES"a um usuário A, com a chave codificadora $(n_A, e_A) = (46927, 39423)$, primeiro temos que encontrar os números equivalentes de "YES", a saber: $24.26^2 + 4.26 + 18 = 16346$, e então calculamos $16346^{39423} \pmod{46927}$, que é $21166 = 1.26^3 + 5.26^2 + 8.26 + 2 =$"BFIC". O receptor A conhece a chave decodificadora $(n_A, d_A) = (46927, 26767)$, e então calcula $21166^{26767} \pmod{46927} = 16346 = 24.26^2 + 4.26 + 18 =$"YES".

A chave do usuário A foi gerada da seguinte forma: primeiro, ele multiplicou os primos $p_A = 281$ e $q_A = 167$ para obter n_A; depois ele escolheu e_A aleatoriamente (mas obedecendo à condição que $mdc(e_A, 280) = mdc(e_A, 166) = 1$). E, então, ele encontrou $d_A = e_A^{-1} \pmod{280.166}$. Os números p_A, q_A e d_A permaneceram secretos. □

Observação 2.3.2. *Nesse exemplo, a etapa que mais consome tempo computacional é a exponenciação modular, por exemplo, $16346^{39423} \pmod{46927}$. Mas se trabalharmos com inteiros muito grandes, provavelmente, a etapa em que o tempo computacional vai ser maior será, para cada usuário A, encontrar dois primos muito grandes p_A e q_A. A fim de escolher rapidamente primos muito grandes, devem ser usados testes eficientes de primalidade.*

Observação 2.3.3. *Ao escolher os primos p_A e q_A, o usuário A deve tomar um cuidado especial. Os dois primos p_A e q_A não podem ser muito próximos (por exemplo, um deles deve*

ter alguns dígitos a mais que o outro). A razão para isso, é que se forem muito próximos, o número $n_A = p_A q_A$ pode ser facilmente fatorado usando a Fatoração de Fermat como segue.

Se $n_A = p_A q_A$ (digamos $p_A > q_A$), então $n_A = \left(\frac{p_A + q_A}{2}\right)^2 - \left(\frac{p_A - q_A}{2}\right)^2$. Se p_A e q_A são muito próximos, então $s = \frac{(p_A - q_A)}{2}$ é pequeno e $t = \frac{(p_A + q_A)}{2}$ é um inteiro apenas ligeiramente maior que $\sqrt{n_A}$ tendo a propriedade que $t^2 - n_A$ é um quadrado perfeito. Se testarmos os inteiros sucessivos $t > \sqrt{n_A}$, logo encontraremos um tal que $n_A = t^2 - s^2$, e assim teremos os fatores primos de n_A, $p_A = (t + s)$ e $q_A = (t - s)$.

Observação 2.3.4. *Vamos ver agora, como enviamos uma assinatura em RSA: Quando discutimos autenticação na seção anterior nós assumimos, para simplificar, que o conjunto de todas as possíveis unidades de mensagem pura e o conjunto de todas as possíveis unidades de mensagem cifrada eram iguais ($\alpha = \beta$). Em RSA é um pouco mais complicado. Aqui, temos que superar o problema de termos diferentes $n'_A s$ e diferentes tamanhos de blocos. Suponha aqui, como na seção anterior, que Alice (A) quer enviar sua assinatura (algum texto puro P) para Bernardo (B). Ela conhece a chave codificadora de Bernardo $K_{E,B} = (n_B, e_B)$ e também conhece sua própria chave decodificadora $K_{D,A} = (n_A, d_A)$. O que ela tem que fazer, é enviar $f_B f_A^{-1}(P)$ se $n_A < n_B$, ou então, $f_A^{-1} f_B(P)$ se $n_A > n_B$. Que é, no primeiro caso, tomar P^{d_A} módulo n_A, depois calcular $(P^{d_A} \pmod{n_A})^{e_B} \pmod{n_B}$, e enviar a Bernardo como um texto cifrado. No caso em que $n_A > n_B$, primeiro ela tem que calcular $P^{e_B} \pmod{n_B}$ e então, trabalhando módulo n_A ela deve elevar esse número a d_A-ésima potência. Claramente, Bernardo vai poder verificar a autenticidade da mensagem no primeiro caso, elevando a d_B-ésima potência módulo n_B e depois a e_A-ésima potência módulo n_A. No segundo caso ele deverá realizar essas duas operação na ordem inversa.*

2.4 Logarítmo discreto

Além do processo usado no sistema RSA - onde se usa o fato de que é fácil multiplicar dois números primos grandes para se obter um número composto n. Mas o processo inverso, onde temos o número composto n e queremos encontrar seus fatores primos, pode ser muito complicado - temos outros processos na teoria dos números, onde também temos essa propriedade do "sentido único". Um deles, é o de se elevar um número a uma potência em um corpo finito grande.

Ao contrário do conjunto dos números reais, onde dados dois números x e b, calcular $log_x b$ é tão fácil quanto calcular b^x, quando estamos trabalhando em um grupo finito, como por exemplo

$\mathbb{F}_q^*$ ou $(\mathbb{Z}/n\mathbb{Z})^*$, podemos calcular com certa facilidade b^x, onde b e x são elementos do grupo, mas dado um número y no grupo, que sabemos ser da forma b^x, encontrar a potência de b que nos dá y, ou em outras palavras, calcular $x = log_b y$ é um problema chamado de "problema do logarítmo discreto. A palavra "discreto"se refere ao fato do grupo ser finito.

Definição 2.4.1. *Dados um grupo G e dois elementos $b, y \in G$ tal que y é uma potência de b, dizemos que o logarítmo discreto de y na base b, é qualquer inteiro x tal que $b^x = y$.*

Exemplo 2.4.1. *Seja $G = \mathbb{F}_{19}^* = (\mathbb{Z}/19\mathbb{Z})^*$ um grupo e seja $b = 2$ um gerador. O logarítmo discreto de 7 na base 2, é 6 ($log_2 7 = 6$)* □

Vamos ver alguns sistemas baseado no problema do logarítmo discreto.

2.4.1 Sistema de troca de chave de Diffie-Hellman

Se compararmos os criptossistemas com chave pública com os clássicos, os que usam chave pública são relativamente mais lentos. Logo, eles podem ser usados de uma forma limitada, juntamente com um sistema clássico, pelo qual as mensagens são transmitidas. Na pŕatica, o processo de incorporar uma chave a um sistema clássico pode ser feito com razoável eficiência usando um sistema de chave pública. Os primeiros a apresentarem uma proposta para isso foram W. Diffie e M.E. Hellman, e eles se basearam no problema do logarítmo discreto.

Vamos descrever agora, o método de Diffie-Hellman para gerar um elemento aleatório de um corpo finito grande $\mathbb{F}_q$. Suponha que q é publicamente conhecido, ou seja, todos sabem qual corpo finito nós estamos usando. Suponha também que g é um elemento fixo de $\mathbb{F}_q$, que também não vai ser mantido secreto. O ideal, é g seja um gerador de $\mathbb{F}_q^*$, no entanto, isso não é necessário. O método que vamos descrever para gerar uma chave, nos dará apena os elementos de $\mathbb{F}_q$ que são potências de g, por isso, se realmente quisermos que os elementos aleatórios de $\mathbb{F}_q^*$ tenham a chance de ser qualquer elemento do corpo, g deve ser um gerador.

Suponha que dois usuários, Alice (A) e Bernardo (B) querem gerar uma chave, que é um elemento aleatório de $\mathbb{F}_q^*$, a qual eles vão usar para codificar suas mensagens. Alice então, escolhe um inteiro aleatório a entre 1 e $q-1$, que ela mantém secreto, e depois calcula $g^a \in \mathbb{F}_q$, e torna público o resultado. Bernardo faz a mesma coisa, ele escolhe um inteiro b e depois torna público g^b. A chave que eles vão usar é, então, g^{ab}. Ambos os usuários podem calcular essa chave. Pois, por exemplo, Alice conhece g^b (que é de conhecimento público) e conhece a. No entanto, uma terceira pessoa conhece apena g^a e g^b.

Se a seguinte suposição vale para o grupo multiplicativo $\mathbb{F}_q^*$, então, uma terceira pessoa não autorizada será incapaz de determinar a chave.

Suposição de Diffie-Hellman: É computacionalmente inviável calcular g^{ab} conhecendo apenas g^a e g^b.

A suposição de Diffie-Hellman é, pelo menos a priori, tão forte quanto a suposição de que o logarítmo discreto não pode ter um cálculo viável no grupo. Isto é, se o logarítmo discreto pode ser calculado então, obviamente, a suposição de Diffie-Hellman falha.

Exemplo 2.4.2. *Suponha que estamos usando uma codificação deslocamento, com letras simples como unidades de mensagem em um alfabeto com 26 letras :* $C \equiv P + B \pmod{26}$. *(aqui estamos usando* B *ao invés de* b *para denotar a chave, de modo que não haja confusão com a letra* b *que usamos nos parágrafos anteriores.) Para escolher* B*, tomamos o menor resíduo não negativo módulo 26 de um elemento aleatório de* $\mathbb{F}_{53}$*. Tome* $g = 2$ *(que é um gerador de* $\mathbb{F}_{53}$*). Suponha que Alice escolha o número aleatório* $a = 29$*, e observa que Bernardo publicou* 2^b*, que é, digamos* $12 \in \mathbb{F}_{53}$*. Ela então pode calcular a chave codificadora, que é* $12^{29} = 21 \in \mathbb{F}_{53}$*, ou seja,* $B = 21$*. Ela por sua vez, torna público* $2^{29} = 45$ *e assim, Bernardo também pode calcular a chave, elevando 45 a b-ésima potência (suponha que Bernardo escolheu* $b = 19$*). É claro que não existe segurança trabalhando com um corpo tão pequeno, pois nesse corpo é fácil encontrar o logarítmo discreto nas base 2, 12 ou 45 módulo 53. Na verdade, em nenhum caso em que se use uma codificação deslocamento, com letras simples como unidades de mensagem se tem segurança.* □

2.4.2 O criptossistema de Massey-Omura

Vamos supor aqui que todos concordam com a escolha do corpo finito $\mathbb{F}_q$, que é fixado e publicado. Então, cada usuário do sistema escolhe secretamente um inteiro aleatório e entre 0 e $q-1$ tal que $mdc(e, q-1) = 1$ e calcula seu inverso $d = e^{-1} \pmod{q-1}$. Se o usuário Alice (A) que enviar uma mensagem P para Bernardo (B), primeiro ela envia para ele o elemento P^{e_A}. Isso não significa nada para Bernardo, já que ele não conhece $e_A^{-1} = d_A$, e assim não pode recuperar P. Então, ele calcula $P^{e_A e_B}$ e manda de volta para Alice. Alice, por sua vez, eleva isso a d_A-ésima potência e obtém P^{e_B} (pois $P^{d_A e_A} = P$), e manda para Bernardo, que assim pode ler a mensagem, pois ele conhece d_B e $P^{d_B e_B} = P$

Observação 2.4.1. *A idéia por trás desse sistema é bem simples. No entanto, é necessário tomar um cuidado especial. Note que é muito importante usar um bom esquema de assinatura*

junto com o sistema Massey-Omura. Caso contrário, qualquer outra pessoa C, como Alice torna público P^{e_A}, pode mandar de volta a Alice a mensagem $P^{e_A e_C}$. Não tendo como saber que essa mensagem não veio de Bernardo, ela eleva isso d_A, publica o resultado e possibilita assim a C ler a mensagem. Então, a mensagem $P^{e_A e_B}$ de Bernardo para Alice deve ser acompanhada por alguma autenticação, ou seja, alguma mensagem, em algum esquema de assinatura que somente Bernardo poderia ter enviado.

2.4.3 Criptossistema de ElGamal

Para começar a descrever esse sistema, vamos fixar um corpo finito bem grande $\mathbb{F}_q$ e um elemento $g \in \mathbb{F}_q^*$ (de preferência, mas não obrigatoriamente, um gerador). Vamos usar unidades de mensagem de texto puro com números equivalentes $P \in \mathbb{F}_q$. Cada usuário A escolhe aleatoriamente um inteiro $a = a_A$, com $0 < a < q - 1$. Esse inteiro a é a chave decodificadora. A chave codificadora (que é de conhecimento público) é o elemento $g^a \in \mathbb{F}_q$.

Para enviar uma mensagem P para o usuário A, nós escolhemos um inteiro aleatório k, e então enviamos a A o seguinte par de elementos de $\mathbb{F}_q$:

$$(g^k, Pg^{ak})$$

Observe que podemos calcular g^{ak} mesmo sem conhecer a, pois conhecemos k (pois nós que escolhemos) e conhecemos g^a, que é a chave codificadora. Agora, o usuário A, que conhece a, pode recuperar P desse par que lhe foi enviado elevando o primeiro elemento a a e depois multiplicando seu inverso pelo segundo elemento.

Capítulo 3

Aplicações

Hoje em dia, as curvas elípticas têm encontrado diversas aplicações. Nesse capítulo vamos discutir algumas delas, como por exemplo, em criptografia.

3.1 Criptossistemas usando curvas elípticas

Já discutimos anteriormente como o grupo abeliano $\mathbb{F}_q^*$, que é o grupo multiplicativo de um corpo finito, pode ser usado para criar um criptossistema com chave pública, baseado na dificuldade para se resolver o problema do logarítmo discreto em corpos finitos. Agora, vamos construir analogamente sistemas com chave pública baseados no grupo abeliano de uma curva elíptica C definida sobre $\mathbb{F}_q$.

3.1.1 Mergulhando o texto puro

Nós queremos, de maneira simples, codificar o texto puro como pontos de uma curva elíptica C definida sobre um corpo finito $\mathbb{F}_q$. Ou seja, queremos que o texto puro m seja prontamente determinado a partir do conhecimento das coordenadas do ponto correspondente P_m. Observe que esse "codificar"não é o mesmo que cifrar. Mais tarde vamos discutir meios para cifrar os pontos P_m (pontos correspondentes ao texto puro).Além disso, um usuário autorizado do sistema deve poder recuperar m depois de ter decifrado os pontos do texto cifrado.

Vamos apresentar agora, um método para mergulhar o texto puro como pontos de um curva elíptica C definida sobre $\mathbb{F}_q$, onde assumimos que $q = p^r$ é grande. Seja k um inteiro. Suponha que nossas unidades de mensagem m são inteiros, tais que $0 \leq m \leq M$. Vamos supor também, que o corpo finito é escolhido de modo que $q > Mk$. Agora, escrevemos os inteiros de 1 até

Mk na forma $mk + j$, onde $1 \leq j \leq k$. Assim, temos uma correspondência injetora entre tais inteiros e o conjunto dos elementos de $\mathbb{F}_q$.

Então, dado m, para cada $j = 1, 2, \ldots, k$ nós obtemos um elemento x de $\mathbb{F}_q$ correspondendo a $mk + j$. Para esse x, nós calculamos o lado direito da equação

$$y^2 = f(x) = x^3 + ax + b,$$

e tentamos encontrar uma raíz quadrada de $f(x)$. Se encontramos um y tal que $y^2 = f(x)$, tomamos $P_m = (x, y)$. Se não $f(x)$ não for um quadrado, nós então trocamos j por $j + 1$ e tentamos com o novo x.

3.1.2 O logarítmo discreto em uma curva

Agora, vamos discutir o problema do logarítmo discreto em um grupo de uma curva elíptica C definida sobre um corpo finito $\mathbb{F}_q$.

Definição 3.1.1. *Seja C uma curva elíptica definida sobre $\mathbb{F}_q$ e B um ponto de C. O problema do logarítmo discreto em C, na base B, é o problema de dado um ponto $P \in C$, encontrar um inteiro $x \in \mathbb{Z}$ tal que $xB = P$, se tal inteiro x existir.*

É provável que o problema do logarítmo discreto em curvas elípticas se mostre mais intratável que o problema em corpos finitos. As técnicas mais fortes desenvolvidas para o uso em corpos finitos parecem não funcionar em curvas elípticas. Isso é especialmente verdade em característica 2. Como pode ser visto em [10], um método para resolver o problema do logarítmo discreto em $\mathbb{F}_{2^r}^*$ torna relativamente fácil calcular logarítmos discretos e, conseqentemente, quebrar os criptossistemas baseados no logarítmo. O método só não funciona quando r é escolhido muito grande. Parece que os sistemas análogos usando curvas elípticas definidas sobre $\mathbb{F}_{2^r}^*$ (veremos mais adiante) são seguros com valores para r significativamente menores. Assim, temos razões práticas para crer que os criptossistemas com chave pública discutidos a seguir (usando curvas elípticas) são mais convenientes nas aplicações do que os sistemas baseados no problema do logarítmo discreto em $\mathbb{F}_q^*$.

Até 1990, os únicos algorítmos que eram conhecidos para tratar do logarítmo discreto em curvas elípticas, eram os que trabalhavam em qualquer grupo, sem levar em consideração sua estrutura particular. Estes algorítmos são exponenciais quando a ordem do grupo tem um grande fator primo, ou seja, é divisível por um primo grande. Mas então, Menezes, Okamoto e Vanstone encontraram uma nova forma de atacar o problema do logarítmo discreto em uma

curva elíptica definida sobre $\mathbb{F}_q$. Eles usaram o par de Weil (ver [12]) para mergulhar o grupo C em grupo multiplicativo de alguma extensão do corpo $\mathbb{F}_q$. Esse mergulho, reduz o problema do logarítmo discreto em C ao problema do logarítmo discreto em $\mathbb{F}_{q^k}^*$.

No entanto, para que a redução ao par de Weil ajude, é essencial que o grau da extensão k seja pequeno. Essencialmente, as únicas curvas elípticas para as quais k é pequeno são chamadas de curvas elípticas supersingulares. Os exemplos mais familiares, são as curvas da forma $y^2 = x^3 + ax$ quando a característica p de $\mathbb{F}_q$ é côngruo a -1 módulo 4, e curvas da forma $y^2 = x^3 + b$ quando p é côngruo a -1 módulo 3. No entanto, a grande maioria das curvas elípticas são não supersingulares. Para elas, a redução quase nunca leva a um algorítmo subexponencial (para mais detalhes ver [7]).

3.1.3 Criptossistema análogo ao sistema com chave trocada de Diffie-Hellman

Suponha que Alice (A) e Bernardo (B) querem encontrar uma chave, que mais tarde, possa ser usada em conjunto com um criptossistema clássico. Primeiro, eles escolhem e publicam um corpo finito $\mathbb{F}_q$ e uma curva elíptica C definida sobre ele. As chaves serão construídas a partir de um ponto aleatório P da curva elíptica. Por exemplo, se eles tem um ponto aleatório $P \in C$, tomando a x-coordenada de P temos um elemento aleatório de $\mathbb{F}_q$, que pode ser convertido em um r-dígito aleatório na base p (onde $q = p^r$) o qual servirá como chave para o criptossistema clássico.

Alice e Bernardo primeiro publicam a escolha de um ponto $B \in C$ para servir como sua "base". B assume o papel do gerador g no sistema Diffie-Hellman para corpos finitos. No entanto, nós não queremos insistir que B seja um gerador do grupo dos pontos da curva C. Na verdade, o grupo pode nem ser cíclico. E mesmo que seja cíclico, queremos evitar o trabalho de verificar se B é um gerador. O que nós queremos, é que o subgrupo gerado por B seja grande, de preferência da mesma ordem que o próprio grupo dos pontos da curva C. Vamos discutir melhor isso mais adiante. Por agora, vamos supor que B é fixado e publicamente conhecido, tomado em C, cuja ordem é muito grande.

Para então gerar uma chave, primeiro Alice escolhe um inteiro aleatório a, que ela mantém secreto. Ela calcula $aB \in C$, e torna público. Bernardo faz a mesma coisa: escolhe um inteiro aleatório b e publica $bB \in C$. A chave secreta que eles vão usar é, então, $P = abB \in C$. Ambos usuários podem calcular essa chave. Por exemplo, Alice conhece bB (que Bernardo tornou público) e conhece a. No entanto, uma terceira pessoa conhece apenas aB e bB. Sem resolver

o problema do logarítmo discreto, ou seja encontrar a conhecendo apenas B e aB, parece não haver maneiras de calcular abB conhecendo apenas aB e bB.

3.1.4 Criptossistema análogo ao sistema Massey-Omura

Este é um criptossistema para transmitir unidades de mensagem m, as quais nós supomos que foram mergulhadas como pontos P_m em alguma curva elíptica (fixada e publicada) C sobre $\mathbb{F}_q$ (onde q é grande). Também supomos que o número N de pontos da curva C foi calculado e publicado. Cada usuário dos sistema escolhe secretamente um inteiro e entre 1 e N tal que $mdc(e, N) = 1$ e calcula seu inverso $d = e^{-1} \pmod{N}$. Se Alice quer enviar a mensagem P_m a Bernardo, primeiro ela envia a ele o ponto $e_A P_m$. Mas como Bernardo não conhece d_A nem e_A, isso não significa nada para ele. Mas, ele multiplica esse ponto por e_B, e envia $e_B e_A P_m$ de volta para Alice. O próximo passo é Alice multiplica o ponto $e_B e_A P_m$ por d_A. Como $NP_m = 0$ e $d_A e_A \equiv 1 \pmod{N}$, isso nos dá o ponto $e_B P_m$, que Alice manda para Bernardo, para que ele possa ler a mensagem multiplicando o ponto $e_B P_m$ por d_B.

3.1.5 Análogo ao sistema ElGamal

Esse é um outro criptossistema com chave pública para transmitir unidades de mensagem P_m. Como no sistema com chave trocada, nós começaremos com um corpo finito $\mathbb{F}_q$ fixado e publicado, uma curva elíptica C definida sobre ele, e um ponto base $B \in C$. (Nesse criptossistema nós não precisamos conhecer o número de pontos N da curva.) Cada usuário escolhe um inteiro aleatório a, que é mantido em segredo. Depois calcula e publica o ponto aB.

Para enviar a mensagem P_m para Bernardo, Alice escolhe um inteiro aleatório k e envia o par de pontos $(kB, P_m + k(a_B B))$. Para ler a mensagem, Bernardo multiplica o primeiro ponto do par pelo inteiro que ele escolheu a_B e subtrai resultado do segundo ponto:

$$P_m + k(a_B B) - a_B(kB) = P_m.$$

3.1.6 A escolha da curva e do ponto

Existem várias maneiras de escolher uma curva elíptica C e um ponto B nela.

- **Escolha aleatória de $(\mathbf{C}, \mathbf{B})$:** Primeiramente, escolha um corpo finito grande $\mathbb{F}_q$. Em seguida, pode-se escolher a curva C e o ponto B ao mesmo tempo da seguinte maneira.

(Vamos assumir que a característica é maior que 3; se $q = 2^r$ ou $q = 3^r$ basta fazer algumas modificações simples.)

Escolha quatro elementos aleatórios x, y, a, b de $\mathbb{F}_q$. Em seguida, tome $c = y^2 - (x^3 + ax^2 + bx)$. Certifique-se de que a cúbica $x^3 + ax^2 + bx + c$ não tem raízes múltiplas. (Se esta condição não é satisfeita, escolha outros pontos aleatórios x, y, a, b.) Seja $B = (x, y)$. Então, B é um ponto da curva elíptica $y^2 = x^3 + ax^2 + bx + c$.

- **Reduzindo $(\mathbf{C}, \mathbf{B})$ módulo p:** Vamos mencionar um segundo caminho para determinar um par consistindo de uma curva elíptica e um ponto nela. Primeiro, escolha curva elíptica "global"e um ponto de ordem infinita nela. Então, seja C uma curva elíptica definida sobre o corpo dos racionais, e seja B um ponto de ordem infinita em C.

  Exemplo 3.1.1. *O ponto $B = (0, 0)$ é um ponto de ordem infinita da curva $C : y^2 + y = x^3 + x^2$.* □

  Em seguida, escolha um primo grande p e considere a redução de C e B módulo p. Mais precisamente, para todo p, exceto para alguns primos pequenos, os coeficientes da equação de C não tem p em seus denominadores, e assim, pode-se considerar os coeficientes da equação módulo p. Se fizermos uma mudança de variável que deixa a equação resultante sobre $\mathbb{F}_q$ na forma $y^2 = x^3 + ax^2 + bx + c$, a cúbica do lado direito da equação não tem raízes múltiplas (exceto no caso de alguns primos pequenos), e então temos uma curva elíptica, que vamos denotar $C \pmod p$, sobre $\mathbb{F}_q$. As coordenadas de B também reduzidas módulo p nos dão um ponto, que será denotado $B \pmod p$, na curva elíptica $C \pmod p$.

  Quando usamos esse método, fixamos C e B, e então podemos gerar muitas curvas diferentes variando o primo p.

- **Ordem do ponto B:** Nesse ponto, é natural surgirem algumas perguntas, como por exemplo, quais as chances de que um ponto aleatório B, escolhido em uma curva aleatória, seja um gerador? Ou, no caso do segundo método para escolher (C, B), quais as chances, já que p varia, de que o ponto B reduzido módulo p seja um gerador de $C \pmod p$? Essa pergunta é análoga a seguinte questão relacionada ao grupo multiplicativo de um corpo finito: Dado um inteiro b, quais as chances, quando p varia, de que b seja um gerador de $\mathbb{F}_q^*$? Essa pergunta foi estudada em ambos os casos: corpo finito e curvas elípticas (para discussão mais detalhada, ver [4]).

Como já dissemos antes, para a segurança dos criptossistemas que mencionamos não é realmente necessário que B seja um gerador. O que é necessário, é que o subgrupo cíclico gerado por B seja um grupo onde o problema do logarítmo discreto seja intratável. Este será o caso quando a ordem de B for divisível por um primo muito grande, digamos, que tem a ordem de magnitude tão grande quanto N.

Um caminho para garantir uma escolha apropriada de B, é escolher a curva elíptica e o corpo finito de forma que o número N de pontos seja ele próprio um número primo. Se fizermos isso, então cada ponto $B \neq 0$ será um gerador. Assim, se usarmos o primeiro método descrito acima, para um corpo fixo $\mathbb{F}_q$, nós continuaríamos escolhendo pares (C, B) até que encontrássemos um em que o número de pontos de C fosse primo. No caso do segundo método, para uma curva elíptica global C sobre $\mathbb{Q}$ nós continuaríamos escolhendo os primos p até que encontrássemos um primo para o qual o número de pontos de $C \pmod{p}$ fosse um número primo.

Observação 3.1.1. *Para que $C \pmod{p}$ tenha possibilidade de ter como número de pontos N um número primo, C deve ser escolhida de forma que tenha torsão trivial, isto é, não tenha pontos de ordem finita, exceto o O. Caso contrário, N será divisível pela ordem do subgrupo de torsão.*

3.2 Fatoração

Uma razão para o interesse em curvas elípticas por parte dos criptógrafos é o uso de curvas elípticas, por H. W. Lenstra, para obter um método de fatoração.

Antes de discutir o algorítmo de fatoração de Lenstra, vamos apresentar uma técnica de fatoração clássica, que é análoga ao método de Lenstra.

Método p − 1 de Pollard: Seja n um número composto, e p algum (ainda desconhecido) fator primo de n. Se p tem a propriedade de que $p - 1$ não tem nenhum divisor primo grande, então o seguinte método certamente encontrará p.

O algorítmo procede da seguinte forma:

1. Escolha um inteiro k que seja múltiplo de todos, ou da maioria, dos inteiros menores do que algum limitante B. Por exemplo, k pode ser $B!$ ou pode ser o mínimo múltiplo comum de todos os inteiros menores ou iguais que B.

2. Escolha um inteiro a entre 2 e $n-2$.

3. Calcule $a^k \pmod{n}$

4. Calcule $d = mdc(a^k - 1, n)$ e o resíduo de $a^k \pmod{n}$ do passo 3.

5. Se d for um divisor trivial de n, comece novamente com uma nova escolha de a e/ou uma nova escolha de k.

Para explicar como o algorítmo funciona, suponha que k é divisível por todos os inteiros positivos menores ou iguais a B, e suponha também que p é um divisor primo de n tal que $p-1$ é um produto de potência de primos pequenos, todos menores que B. Então, temos que k é um múltiplo de $p-1$, pois é um múltiplo de todas as potências de primos da fatoração de $p-1$, e assim, pelo Pequeno Teorema de Fermat, temos que $a^k \equiv 1 \pmod{p}$. Então, p divide $mdc(a^k - 1, n)$, e assim, a única maneira de termos um fator não trivial no passo 4 é se acontecer $a^k \equiv 1 \pmod{n}$.

Exemplo 3.2.1. *Queremos fatorar* $n = 540143$ *usando o método descrito acima. Tome* $B = 8$, $k = 840$ *(que é o mínimo múltiplo comum entre* $1, 2, \ldots, 8$*) e* $a = 2$. *Fazendo as contas, temos que* $a^k = 2^{840} = 54047 \pmod{n}$ *e* $d = mdc(a^k - 1, n) = mdc(2^{840} - 1, 540143) = mdc(53046, 540143) = 421$. *Isso nos dá a fatoração* $540143 = 421.1283$. □

A fraqueza do método de Pollard é clara, falha se tentarmos usá-lo quando para todo divisor primo p de n o número $p-1$ tem em sua fatoração apenas primos relativamente grandes.

Exemplo 3.2.2. *Seja* $n = 491389$. *É improvável que encontremos um divisor não trivial enquanto não escolhermos* B *maior ou igual a 191. Isso porque temos que* $n = 383.1283$, $383 - 1 = 2.191$, $1283 - 1 = 2.641$, *e* 191 *e* 641 *são primos. Exceto para* $a \equiv 0, \pm 1 \pmod{383}$, *todos os outros* $a's$ *têm ordem módulo 383 igual a 191 ou 382; e exceto para* $a \equiv 0, \pm 1 \pmod{1283}$, *todos os outros* $a's$ *têm ordem módulo 1283 igual a 641 ou 1282. Assim, a menos que* k *seja divisível por 191 (ou 641), é provável que encontremos repetidas vezes que* $mdc(a^k - 1, n) = 1$ *no passo 4.* □

O dilema básico com o método $p-1$ de Pollard é que fixamos nossas esperanças no grupo $(\mathbb{Z}/p\mathbb{Z})^*$, mais precisamente, os vários grupos onde p é um divisor primo de n. Para um n fixo, esses grupos são fixos. E se todos eles tiverem a ordem divisível por um primo grande, o método estará comprometido.

A grande diferença no método de Lenstra, como vamos ver mais adiante, é que trabalhando com curvas elípticas sobre $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$, nós teremos uma quantidade de grupos muito maior

que poderemos usar e, assim, poderemos realmente esperar encontrar sempre um grupo cuja ordem não é divisível por um primo grande.

Antes de descrevermos o algorítmo de Lenstra, faremos alguns comentários sobre reduzir módulo n os pontos da curva elíptica, onde n é um número composto.

Redução módulo n: No decorrer dessa seção, n denotará um inteiro composto ímpar e p (ainda desconhecido) um fator primo de n. Vamos supor também que $p > 3$. Para qualquer inteiro m e quaisquer dois racionais x_1 e x_2 com denominadores primos com m, escrevemos $x_1 \equiv x_2 \pmod{m}$, se $x_1 - x_2$ for uma fração com numerador divisível por m. Para qualquer número racional x_1 com denominador primo com m existe um único inteiro x_2, chamado de menor resíduo não negativo, entre 0 e $m-1$ tal que $x_1 \equiv x_2 \pmod{m}$. As vezes denotamos esse menor resíduo não negativo por $x_1 \pmod{m}$.

Suponha que temos uma equação da forma $y^2 = x^3 + ax^2 + bx + c$ com $a, b, c \in \mathbb{Z}$, e um ponto $P = (x, y)$ que a satisfaça. Na prática, a curva C junto com o ponto P podem ser gerados de alguma maneira aleatória, como por exemplo escolhendo quatro inteiros a, b, x, y e então fazendo $c = y^2 - x^3 - ax^2 - bx$. Vamos assumir que a cúbica tem raízes distintas, isto é, $D = -4a^3c + a^2b^2 + 18abc - 4b^3 - 27c^2 \neq 0$ (o discriminante é diferente de zero). Para simplificar, vamos supor também que o discriminante $D = -4a^3c + a^2b^2 + 18abc - 4b^3 - 27c^2$ não tem fator comum com n; em outras palavras, $x^3 + ax^2 + bx + c$ não tem raízes múltiplas módulo p para qualquer divisor primo p de n. Na prática, para cada escolha de a, b and c, nós podemos verificar isto calculando $d = mdc(D, n)$. Se $d > 1$, então ou n divide D, e nesse caso teremos que fazer uma nova escolha de a, b and c, ou então obtivemos um divisor não trivial de n. Por isso, vamos supor sempre que $mdc(D, n) = 1$.

Assim, vamos considerar o ponto P e todos os seus múltiplos módulo n. Isto significa que tomamos $P \pmod{n} = (x \pmod{n}, y \pmod{n})$, e , cada vez que calculamos algum múltiplo kP, nós realmente calculamos apenas a redução módulo n das coordenadas. Para trabalharmos módulo n há uma condição não trivial que deve ser satisfeita sempre que multiplicamos ou somamos dois pontos diferentes. A saber, todos os denominadores devem ser primos com n.

Proposição 3.2.1. *Seja C uma curva elíptica com equação $y^2 = x^3 + ax^2 + bx + c$, onde $a, b, c \in \mathbb{Z}$ e $mdc(D, n) = 1$, onde D denota o discriminante da curva. Sejam P_1 e P_2 dois pontos de C, cujas coordenadas têm denominadores primos com n e $P_1 \neq -P_2$. Então, $P_1 + P_2 \in C$ tem coordenadas com denominadores primos com n se, e somente se, não existe um divisor primo p de n com a seguinte propriedade: os pontos $P_1 \pmod{p}$ e $P_2 \pmod{p}$ da curva elíptica $C \pmod{p}$ quando somados resultam no ponto no infinito $\mathcal{O} \pmod{p} \in C \pmod{p}$. Onde C*

(mod p) *denota a curva elíptica sobre* $\mathbb{F}_p$ *obtida reduzindo módulo* p *os coeficientes da equação* $y^2 = x^3 + ax^2 + bx + c$.

Demonstração: Primeiro, suponha que $P_1 = (x_1, y_1), P_2 = (x_2, y_2)$ e $P_1 + P_2 \in C$ têm coordenadas cujos denominadores são primos com n. Seja p qualquer número primo, divisor de n. Vamos mostrar que $P_1 \pmod p + P_2 \pmod p \neq \mathcal{O} \pmod p$.

- Se $x_1 \not\equiv x_2 \pmod p$, então, concluimos que

$P_1 \pmod p + P_2 \pmod p$ não é o ponto no infinito em $C \pmod p$.

- Agora, suponha que $x_1 \equiv x_2 \pmod p$.

Primeiro, se $P_1 = P_2$, então as coordenadas de $P_1 + P_2 = 2P_1$ são dadas pela fórmula (pela proposição 1.4.1)

$$2P_1 = (\lambda^2 - a - 2x_1, -\lambda(\lambda^2 - a - 2x_1) - v),$$

onde

$$\lambda = \frac{3x_1^2 + 2ax_1 + b}{2y_1} \quad \text{e} \quad v = \frac{-x_1^3 + bx_1 + 2c}{2y_1}.$$

e $2P_1 \pmod p$ é dado pela mesma fórmula com cada termo substituído pelo seu resíduo módulo p. Devemos mostrar que o denominador $2y_1$ não é divisível por p. Suponha que seja divisível. Então, como o denominador da x-coordenada de $2P_1$ não é divisível por p, segue que o numerador é divisível por p. Mas isso significa que x_1 é uma raiz módulo p da cúbica $x^3 + ax^2 + bx + c$ e da sua derivada, contradizendo a hipótese de que não existem raízes múltiplas módulo p.

Agora suponha que $P_1 \neq P_2$. Como $x_1 \equiv x_2 \pmod p$ e $x_2 \neq x_1$ podemos escrever $x_2 = x_1 + p^r x$ com $r \geq 1$ escolhido tal que nem o numerador nem o denominador de x sejam divisíveis por p. Como assumimos que $P_1 + P_2$ tem denominador não divisível por p, podemos usar a fórmula (dada pela proposição 1.4.1)

$$x_{(P_1+P_2)} = \lambda^2 - a - x_1 - x_2$$

$$y_{(P_1+P_2)} = -\lambda(\lambda^2 - a - x_1 - x_2) - v,$$

onde

$$\lambda = \frac{y_2 - y_1}{x_2 - x_1} \quad \text{e} \quad v = \frac{y_1 x_2 - y_2 x_1}{x_2 - x_1}.$$

para concluir que y_2 é da forma $y_1 + p^r y$.

Por outro lado,

$$y_2^2 = (x_1+p^r x)^3+a(x_1+p^r x)^2+b(x_1+p^r x)+c \equiv x_1^3+ax_1^2+bx_1+c+p^r x(3x_1^2+2ax_1+b) = \quad (3.1)$$

$$= y_1^2 + p^r x(3x_1^2 + 2ax_1 + b) \pmod{p^{r+1}}$$

Mas como $x_2 \equiv x_1 \pmod p$ e $y_2 \equiv y_1 \pmod p$ segue que $P_1 \pmod p \equiv P_2 \pmod p$ e então $P_1 \pmod p + P_2 \pmod p = 2P_1 \pmod p$, que é $\mathcal{O} \pmod p$ se, e somente se, $y_2 \equiv y_1 \equiv 0 \pmod p$..

Se essa congruência ocorrer, temos que $y_2^2 - y_1^2 = (y_2 - y_1)(y_2 + y_1)$ é divisível por p^{r+1} (isso é, o numerador é) e assim, 3.1 implica que $3x_1^2 + 2ax_1 + b \equiv 0 \pmod p$. O que não ocorre pois $x^3 + ax^2 + bx + c \pmod p$ não tem raízes múltiplas e então x_1 não pode ser raiz do polinômio e da derivada módulo p. Logo, concluímos que $P_1 \pmod p + P_2 \pmod p \neq \mathcal{O} \pmod p$.

Reciprocamente, suponha que para qualquer divisor p de n, temos $P_1 \pmod p + P_2 \pmod p \neq \mathcal{O} \pmod p$. Queremos provar que as coordenadas de $P_1 + P_2$ têm denominadores primos com n, isto é, que os denominadores não são divisíveis por p, para qualquer p que divide n.

Fixe alguma p, tal que p divide n.

- Se $x_2 \not\equiv x_1 \pmod p$ então por 1.4.1 temos que não existe denominador divisível por p.
- Suponha $x_2 \equiv x_1 \pmod p$.

Então, $y_2 \equiv \pm y_1 \pmod p$, mas como $P_1 \pmod p + P_2 \pmod p \neq \mathcal{O} \pmod p$ devemos ter $y_2 \equiv y_1 \not\equiv \mathcal{O} \pmod p$.

Primeiro, se $P_2 = P_1$, por 1.4.1 e pelo fato de que $y_1 \not\equiv \mathcal{O} \pmod p$ temos que as coordenadas de $P_1 + P_2 = 2P_1$ tem denominador primo com p.

Se $P_1 \neq P_2$ escreva $x_2 = x_1 + p^r x$ com x não divisível por p e use 3.1 para escrever

$$\frac{(y_2^2 - y_1^2)}{(x_2 - x_1)} = 3x_1^2 + 2ax_1 + b \pmod p$$

Como p não divide $y_2 + y_1 \equiv 2y_1 \pmod p$ segue que não existe p no denominador de $\dfrac{(y_2^2 - y_1^2)}{(y_2 - y_1)(x_2 - x_1)} = \dfrac{y_2 - y_1}{x_2 - x_1}$ e então pela proposição 1.4.1 não existe p nas coordenadas de $P_1 + P_2$. □

Método de Lenstra: Seja n um inteiro composto. Queremos encontrar um fator não trivial d de n, ou seja, encontrar um número tal que d divide n e $1 < d < n$. Vamos começar tomando alguma curva elíptica $C : y^2 = x^3 + ax^2 + bx + c$ com coeficientes inteiros, e um ponto $P = (x, y)$ nessa curva. O par (C, P) é gerado de alguma maneira aleatória. Tentaremos usar C e P para fatorar n, como será explicado mais adiante; mas se a tentativa falhar, escolhemos outro par (C, P), e tentamos novamente. Continuamos com esse processo, até que seja encontrado um divisor d não trivial de n. Se a probabilidade de o método falhar é $\rho < 1$, então a probabilidade de que h sucessivas escolhas dos pares (C, P) falhem é ρ^h, que é pequena para h grande. Assim, com uma probabilidade elevada, vamos conseguir fatorar n com um número razoável de tentativas.

Uma vez que temos um par (C, P), escolhemos um inteiro k que é divisível por potências de primos pequenos (tais primos são menores que um certo limite B) que são menores que algum limitante L. Isto é, tomamos

$$k = \prod_{\ell \leq B} \ell^{\alpha_\ell}, \tag{3.2}$$

onde $\alpha_\ell = [log\, L \setminus log\, \ell]$ é o maior expoente tal que $\ell^{\alpha_\ell} \leq L$. Nós então, tentamos calcular kP, trabalhando todo o tempo módulo n. Este cálculo é simples, a menos que encontremos a seguinte dificuldade: quando tentamos encontrar o inverso de $2y$ (necessário para calcular as coordenadas de $2P$), ou o inverso de $x_2 - x_1$ (necessário para calcular as coordenadas de $P_1 + P_2$) encontramos um número que não é primo com n. De acordo com a proposição 3.2.1 isso ocorre quando temos algum múltiplo $k_1 P$ (uma soma parcial encontrada ao longo do caminho para calcular kP) que para algum p que divide n tem a propriedade $k_1 P \pmod p = \mathcal{O} \pmod p$, isto é, o ponto $P \pmod p$ no grupo de $C \pmod p$ tem uma ordem que divide k_1.

Usando o algorítmo de Euclides para tentar encontrar o inverso módulo n para um denominador que é divisível por p, nós encontramos, ao invés disso, o máximo divisor comum entre n e esse denominador. Esse máximo divisor comum terá a propriedade de ser um divisor de n, a menos que seja o próprio n, ou seja, a menos que o denominador seja divisível por n. Isso significa , pela proposição 3.2.1 que $k_1 P \pmod p = \mathcal{O} \pmod p$ para todo divisor primo p de n. Então, é certo que quando tentamos calcular $k_1 P$ módulo n para um k_1 que é um múltiplo da ordem de $P \pmod p$ para algum divisor p de n, nós vamos obter um divisor próprio para n.

Observação 3.2.1. *Note que é semelhante ao método $p-1$ de Pollard. Mas ao invés do grupo $(\mathbb{Z}/p\mathbb{Z})^*$, estamos usando o grupo $C \pmod p$. No entanto, se C for uma escolha ruim, isto*

é, para cada divisor p de n o grupo $C \pmod p$ tem ordem divisível por um primo grande (e então, $kP \pmod p$ não é igual a $\mathcal{O} \pmod p$) nós temos maneiras de tomar outra curva C junto com um ponto $P \in C$. E nós não tínhamos essa opção com o método $p-1$ de Pollard.

O algorítmo: Seja n um inteiro positivo ímpar. Vamos descrever o método de Lenstra.

Suponha que temos um método para gerar pares (C, P), consistindo de uma curva elíptica $y^2 = x^3 + ax^2 + bx + c$, com $a, b, c \in \mathbb{Z}$ e um ponto $P = (x, y) \in C$. Tendo um par, nós seguimos o procedimento descrito a seguir. Se o procedimento falhar, e não obtivermos um fator não trivial de n, então geramos outro par (C, P) e repetimos o processo.

Antes de trabalhar com $C \pmod n$, devemos verificar se é de fato uma curva elíptica módulo p, onde p é qualquer divisor de n, isto é, a cúbica tem que ter raízes distintas módulo p. Isso acontece se, e somente se, o discriminante D é primo com n. Assim, se $mdc(D, n) = 1$, podemos prosseguir. E claro que se o mdc é estritamente maior que 1 e estritamente menor que n, já temos um divisor de n, e aí acabou. Se o mdc é igual a n, então temos que escolher uma curva elíptica diferente.

Em seguida, suponha que escolhemos dois inteiros limitantes B e L. Aqui, B é um limitante para os divisores primos do inteiro k pelo qual vamos multiplicar o ponto P. Já o número C, aproximadamente falando, é um limitante para os divisores primos p de n para os quais é bem provável que satisfaça a relação $kP \pmod p = O \pmod p$.. Nós então escolhemos k, que é dado pela fórmula 3.2, ou seja, k é o produto de todas as potências menores ou iguais a L, as quais são potências de primos menores ou iguais a B. Então, teorema de Hasse nos diz que, se p é tal que $p + 1 + 2\sqrt{p} < L$ e a ordem de $C \pmod p$ não é divisível por nenhum primo maior que B, então k é um múltiplo dessa ordem e portanto $kP \pmod p = O \pmod p$.

Agora, trabalhando módulo n, tentamos calcular kP da seguinte forma. Calcule $2P$, $2(2P)$, $2(4P)$, ..., $2^{\alpha_2}P$, depois $3(2^{\alpha_2})P$, $3(3.2^{\alpha_2}P)$, ..., $3^{\alpha_3}2^{\alpha_2}P$, e assim por diante até que tenhamos $\prod_{\ell \leq B} \ell^{\alpha_\ell} P$. Nesses cálculos, sempre que temos que dividir módulo n, usamos o algorítmo de Euclides para encontrar o inverso módulo n. Se em qualquer estágio do algorítmo de Euclides não encontrarmos um inverso, teremos encontrado um divisor não trivial de n, ou teremos encontrado o próprio n como mdc entre n e o denominador. No primeiro caso, o algorítmo obteve sucesso. No segundo caso, devemos voltar e escolher um novo par (C, P). Se o algorítmo de Euclides sempre fornecer um inverso - e então kP módulo n é calculado com sucesso - então também devemos voltar e escolher um novo par (C, P). Essa é a descrição do algorítmo do método de Lenstra.

Exemplo 3.2.3. *Seja uma família de curvas elípticas* $y^2 = x^3 + ax - a$, $a = 1, 2 \ldots$, *cada qual contendo o ponto* $P = (1, 1)$. *Antes de usar um a para um dado n, devemos verificar que o discriminante* $4a^3 + 27a^2$ *é primo com n. Vamos tentar fatorar o número* $n = 5429$ *com* $B = 3$ *e* $L = 92$. *(Nesse exemplo, vamos mostrar como o método funciona usando um valor pequeno para n. claro que na prática, o método torna-se valioso apenas para valores de n muito maiores.). Aqui, a escolha de L foi motivada por nosso desejo de encontrar um fator primo p quase tão grande quanto* $\sqrt{n} \approx 73$*; para* $p = 73$ *o limite do número de* $\mathbb{F}_p$ *pontos em uma curva elíptica é, de acordo com o teorema de Hasse,* $74 + 2\sqrt{p} < 92$. *Usando a fórmula* 3.2, *nós escolhemos* $k = 2^6.3^4$. *Para cada valor de a, nós sucessivamente multiplicamos P por 2 seis vezes e em seguida multiplicamos por 3 quatro vezes, trabalhando módulo n, na curva elíptica* $y^2 = x^3 + ax - a$. *Quando* $a = 1$ *nós observamos que a multiplicação procede suavemente, e que* $3^4.2^6 \pmod{p}$ *é um ponto finito em* $C \pmod{p}$ *para todos divisores p de n. Então, tentamos com* $a = 2$. *Observamos que quando tentamos calcular* $3^2.2^6P$, *nós obtemos um denominador cujo mdc com n é o fator 61. Isto é, a ordem do ponto* $(1, 1)$ *divide* $3^2.2^6$ *na curva* $y^2 = x^3 + 2x - 2$ *módulo 61. Então, na segunda tentativa obtivemos sucesso. Por outro lado, se usarmos* $a = 3$, *esse método nos dará outro fator primo, a saber 89, quando tentarmos calcular* $3^4.2^6$. □

3.3 Números congruentes

Um número inteiro $n \geq 1$ é dito um número congruente se existir um triângulo retângulo cujos lados sejam números racionais e cuja área seja n.

A primeira etapa para determinar se um número n é congruente é obter números racionais x, ye z que sejam lados de um triângulo retângulo, ou seja tais que $x^2 + y^2 = z^2$. Uma terna de inteiros positivos (x, y, z) tais que $x^2 + y^2 = z^2$ é chamada terna pitagórica. Para encontrar x, y e z tomamos inteiros $a > b > 0$ e traçamos no plano XY a reta passando por $(-1, 0)$ com coeficiente angular b/a. Esta reta interceptacta o círculo $S^1 = \{(X, Y) \in \mathbb{R}^2 \,;\, X^2 + Y^2 = 1\}$ no ponto

$$u = \frac{a^2 - b^2}{a^2 + b^2} \qquad \text{e} \qquad v = \frac{2ab}{a^2 + b^2}$$

gerando a terna

$$x = a^2 - b^2,\ y = 2ab,\ z = a^2 + b^2 \tag{3.3}$$

Assim, o problema de achar números congruentes se reduz a achar ternas pitagóricas tais que $n = xy/2$.

Antes de estendermos a noção de números congruentes a números racionais positivos, lembremos que um inteiro positivo n é dito livre de quadrados, se n puder ser escrito da forma $n = \prod_{i=1}^{r} p_i$, onde os $p_i's$ são números primos distintos.

Um número racional positivo n é um número congruente se existirem $x, y, z \in \mathbb{Q}$ positivos tais que $x^2 + y^2 = z^2$ e $n = xy/2$. Notemos que existe $s \in \mathbb{Q} - \{0\}$ tal que $s^2 n \in \mathbb{Z}$ é livre de quadrados. A área do triângulo retângulo de lados sx, sy e sz é igual a $s^2 n$. Ou seja, $s^2 n$ é um número inteiro congruente. Assim podemos sempre supor que n seja um inteiro positivo livre de quadrados. Seja $(\mathbb{Q}^+)^2 = \{x^2\,;\, x \in \mathbb{Q}^+ \,\mathrm{e}\, x \neq 0\}$ o grupo multiplicativo dos números racionais positivos que são quadrados. O argumento acima mostra que o fato de n ser um número congruente depende de sua classe módulo $(\mathbb{Q}^+)^2$ e que nesta classe sempre existe um número inteiro $m \geq 1$ livre de quadrados.

Uma primeira "receita ingênua"para produzir um número congruente n é utilizar 3.3 para listar todas as possíveis ternas pitagóricas. Depois para cada terna calcular a respectiva área e verificar se n ocorre na lista das áreas. É claro que este procedimento está longe de fornecer um método eficiente.

Podemos caracterizar um número congruente da seguinte forma.

Proposição 3.3.1. *Seja $n \geq 1$ um número inteiro livre de quadrados. Sejam $x, y, z \in \mathbb{Q}$ tais que $0 < x < y < z$. Então existe uma bijeção entre o conjunto dos triângulos retângulos de lados x, y, z e área n e o conjunto de números racionais w tais que*

$$w, \quad w + n, \quad w - n \in (\mathbb{Q}^+)^2$$

dada por

$$(x, y, z) \longmapsto w = \left(\frac{z}{2}\right)^2$$

com inversa

$$w \longmapsto (\sqrt{w+n} - \sqrt{w-n}, \sqrt{w+n} + \sqrt{w-n}, 2\sqrt{w})$$

Em particular, n é um número congruente se, e somente se, existir um número racional w tal que

$$w,\ w + n,\ w - n \in (\mathbb{Q}^+)^2.$$

Demonstração: Se $x^2 + y^2 = z^2$ e $n = \dfrac{xy}{2}$, então $(x \pm y)^2 = z^2 \pm 4n$. Logo

$$\left(\frac{x \pm y}{2}\right)^2 = \left(\frac{z}{2}\right)^2 \pm n.$$

Tomando $w = \left(\dfrac{z}{2}\right)^2$ temos que $w,\ w+n,\ w-n \in (\mathbb{Q}^+)^2$.

Reciprocamente, dado w tal que $w,\ w+n,\ w-n \in (\mathbb{Q}^+)^2$ então $x = \sqrt{w+n} - \sqrt{w-n}$, $y = \sqrt{w+n} + \sqrt{w-n}$ e $z = 2\sqrt{w}$ satisfazem a $0 < x < y < z$, $xy = 2n$ e $x^2 + y^2 = z^2$. □

3.3.1 Equações Cúbicas

Nesta seção mostraremos como associar a um número congruente n uma solução de uma certa equação cúbica, demonstrando que um número é congruente se, e somente se, o grupo $C_{F_n}(\mathbb{Q})$ tem elementos de ordem infinita.

Para que isso possa ser feito, vamos inicialmente apresentar algumas definições e resultados.

Seja n um número congruente e $x, y, z \in \mathbb{Q}$ tais que $0 < x < y < z$, $n = \dfrac{xy}{2}$ e $x^2 + y^2 = z^2$. Temos que

$$\left(\frac{x^2 - y^2}{4}\right)^2 = \left(\frac{z}{2}\right)^4 - n^2.$$

Em outras palavras, encontramos soluções racionais $u = \dfrac{z}{2}$ e $v = \dfrac{x^2 - y^2}{4}$ para a equação $u^4 - n^2 = v^2$. Multiplicando ambos os membros desta igualdade por u^2 obtemos

$$(u^2)^3 - n^2u^2 = (uv)^2.$$

Portanto, $a = u^2$ e $b = uv$ fornece uma solução racional (a, b) para a equação cúbica $Y^2 = X^3 - n^2X$.

Reciprocamente, dada uma solução racional (a, b) da equação cúbica $Y^2 = X^3 - n^2X$, perguntamos se (a, b) provém de um triângulo retângulo como acima. Isto nem sempre é verdade. Primeiro é necessário que $a \in (\mathbb{Q}^+)^2$. Além disso o denominador de a tem que ser par. De fato, dada uma terna pitagórica $x < y < z$, seja s o *mmc* dos denominadores de x, y e z. Logo, $x' = sx, y' = sy, z' = sz$ são números inteiros primos entre si. Nesse caso, x' e y' têm paridades distintas, digamos que x' seja ímpar e y' seja par. Em particular z' é ímpar. Portanto, $a = \left(\dfrac{z}{2}\right)^2 = \left(\dfrac{z'}{2s}\right)^2$ tem denominador par.

Proposição 3.3.2. *Seja $(a, b) \in \mathbb{Q} \times \mathbb{Q}$ uma solução de $Y^2 = X^3 - n^2X$ tal que*

$$a \in (\mathbb{Q}^+)^2 \text{ com denominador par.}$$

Então, existe um triângulo retângulo de área n e lados $\sqrt{a+n}-\sqrt{a-n}$, $\sqrt{a+n}+\sqrt{a-n}$ e $2\sqrt{a}$.

Demonstração: Seja $u=\sqrt{a}\in\mathbb{Q}, u>0$ e $v=\dfrac{b}{u}$. Então, $v^2=\dfrac{b^2}{a}=a^2-n^2$. Seja t o denominador de u. Logo, os denominadores de v^2 e a^2 são iguais a t^4, em particular

$$(t^2v, t^2n, t^2a)$$

é uma terna pitagórica com t^2n par e $mdc(t^2v, t^2n, t^2a)=1$.

Uma terna pitagórica x, y, z tal que $mdc(x, y, z)=1$ é chamada de uma terna pitagórica primitiva. Suponhamos que y seja par, logo x e z são ímpares. Sejam $A, B, C>0$ números inteiros tais que $y=2C$, $z+x=2A$ e $z-x=2b$. Observemos que $mdc(A, B)=1$ e $AB=C^2$. Logo existem inteiros positivos α e β tais que $A=\alpha^2$ e $B=\beta^2$. Em particular, $z=A+B=\alpha^2+\beta^2$, $x=A-B=\alpha^2-\beta^2$ e $y=z^2-x^2=(2\alpha\beta)^2$. Portanto, $y=2\alpha\beta$.

Aplicando esse argumento à terna (t^2v, t^2n, t^2a) obtemos que existem positivos α, β tais que,

$$t^2v=\alpha^2-\beta^2,\quad t^2n=2\alpha\beta,\quad t^2a=\alpha^2+\beta^2$$

Consequentemente o triângulo retângulo de lados

$$\left(\frac{2\alpha}{t}, \frac{2\beta}{t}, 2u\right)$$

tem área $\dfrac{2\alpha\beta}{t^2}=n$. Pela proposição 3.3.2 temos que esta terna corresponde a $\left(\dfrac{2u}{2}\right)^2=u^2=a$. Logo existe um triângulo retângulo de lados $\sqrt{a+n}-\sqrt{a-n}$, $\sqrt{a+n}+\sqrt{a-n}$ e $2\sqrt{a}$ de área n. □

Observação 3.3.1. *A equação cúbica obtida no parágrafo anterior é um exemplo de uma curva elíptica.*

Definição 3.3.1. *Seja C uma curva elíptica. Definimos e denotamos o conjunto* $C(\mathbb{Q})=\{(a:b:c)\in C\,;\, a, b, c\in\mathbb{Q}\}$

O próximo teorema é muito importante na teoria das curvas elípticas e pode ser visto com detalhes em [13]

Teorema 3.3.1 (Teorema de Mordell-Weil)**.** *$C(\mathbb{Q})$ é um grupo abeliano finitamente gerado.*

Observação 3.3.2. *Seja $C(\mathbb{Q})_{tor}$ o subgrupo de $C(\mathbb{Q})$ dos elementos de ordem finita. Pelos Teoremas de Mordell-Weill e da Decomposição de Grupos Abelianos Finitamente Gerados, que pode ser visto em [3], segue que existe um isomorfismo de grupos*

$$C(\mathbb{Q}) \cong C(\mathbb{Q})_{tor} \oplus \mathbb{Z}^r,$$

onde $C(\mathbb{Q})_{tor}$ é finito.

Dizemos que r é o posto algébrico de C.

Para podermos atingir nosso objetivo principal de relacionar um número congruente a uma solução de uma equação cúbica, é necessário caracterizarmos os pontos de ordem 2.

Seja $P = (a : b : 1), Q = (a : -b : 1) \in C$ e $l \subset \mathbb{P}^2_{\mathbb{C}}$ a reta passando por P e Q. Observemos que, neste caso, $\mathcal{O}$ é o terceiro ponto de $l \cap C$. Consideremos a reta

$\ell = \{(a' : b' : c') \,;\, c' = 0\}$

Esta é a reta tangente a C em $\mathcal{O}$. Logo, concluimos que $Q = -P$.

Um ponto P é de ordem 2 se, e somente se, $P = -P$. Isto significa que $b = 0$. Se γ_1, γ_2 e γ_3 são as 3 raízes distintas de $X^3 + AX + B$, então os pontos de ordem 2 de C são

$$(\gamma_1 : 0 : 1), \quad (\gamma_2 : 0 : 1) \quad (\gamma_3 : 0 : 1).$$

Também é necessário discutirmos um pouco como o grupo de uma curva elíptica se comporta quando reduzimos módulo p. Seja p um número primo, $\mathbb{F}_{\mathrm{p}}$ o corpo finito de p elementos, $\mathbb{P}^2_{\mathbb{F}_{\mathrm{p}}}$ e $\mathbb{P}^2_{\mathbb{Q}}$ os planos projetivos definidos sobre $\mathbb{F}_{\mathrm{p}}$ e $\mathbb{Q}$, respectivamente. Dado $(a : b : c) \in \mathbb{P}^2_{\mathbb{Q}}$ podemos sempre escolher representantes $a_0, b_0, c_0 \in \mathbb{Z}$. Para isso basta multiplicar a, b, c pelo mínimo múltiplo comum dos denominadores, por exemplo. Além disso, podemos fazer esta escolha de tal forma que $mdc(a_0, b_0, c_0) = 1$. Assim, definimos a aplicação

$$\begin{array}{rcl} \Phi \,:\, \mathbb{P}^2_{\mathbb{Q}} & \longrightarrow & \mathbb{P}^2_{\mathbb{F}_{\mathrm{p}}} \\ P = (a_0 : b_0 : c_0) & \longrightarrow & \tilde{P} = (\tilde{a_0} : b_0 \tilde{:} \tilde{c_0}) \end{array}$$

Proposição 3.3.3. *Seja $i \in \{1, 2\}$ e $P_i = (x_i : y_i : z_i) \in \mathbb{P}^2_{\mathbb{Q}}$. A igualdade $\Phi(P_1) = \Phi(P_2)$ ocorre se, e somente se, p dividir simultaneamente os números $(y_1z_2 - y_2z_1), (x_2z_1 - x_1z_2)$ e $(x_1y_2 - x_2y_1)$.*

Demonstração: Observemos que $\Phi(P_1) = \Phi(P_2)$ ocorre se, e somente se, os vetores $(\tilde{x_1}, \tilde{y_1}, \tilde{z_1})$ e $(\tilde{x_2}, \tilde{y_2}, \tilde{z_2})$ são $\mathbb{F}_{\mathrm{p}}$-linearmente dependentes, o que equivale à condição acima. □

Agora, sejam $n \geq 1$ um número inteiro, e a curva elíptica $C_{F_n} : F_n(X, Y, Z) = Y^2Z - X^3 + n^2XZ^2 \in \mathbb{Z}[X, Y, Z]$. Seu discriminante Δ_n é igual a $4n^6$.

Seja $p > 2$ um número primo e $C_{\tilde{F}_n} : \tilde{F}_n(X,Y,Z) = Y^2Z - X^3 + \tilde{n}^2Z^2 \in \mathbb{F}_p[X,Y,Z]$ a redução de C_{F_n} módulo p. Para que $C_{\tilde{F}_n}$ defina uma curva elíptica sobre $\mathbb{F}_p$ é necessário e suficiente que $4\tilde{n}^6 \neq \tilde{0}$ em $\mathbb{F}_p$. Isto é satisfeito se p não divide n e $p > 2$ (o que estamos supondo). Seja $C_{F_n}(\mathbb{F}_p) = \{(\tilde{a_0} : \tilde{b_0} : \tilde{c_0}) \in \mathbb{P}^2_{\mathbb{F}_p}\,;\, (a_0 : b_0 : c_0) \in C(\mathbb{Q})\}$.

Neste caso, Φ induz uma aplicação

$$\Phi_n : C_{F_n}(\mathbb{Q}) \longrightarrow C_{\tilde{F}_n}(\mathbb{F}_p)$$

A proposição 1.4.1 garante que a adição em C_{F_n} preserva $C_{F_n}(\mathbb{Q})$. A partir das fórmula da proposição 1.4.1 podemos definir a adição em $C_{\tilde{F}_n}$. Novamente, esta adição preserva $C_{\tilde{F}_n}(\mathbb{F}_p)$. Além disso, como $p > 2$ então Φ_n é um homomorfismo de grupos.

Proposição 3.3.4. *Se* $p \equiv 3 \pmod 4$ *então* $\#C_{\tilde{F}_n}(\mathbb{F}_p) = p + 1$.

Demonstração: Notemos que $(0:0:1), (\tilde{n}:0:1), (-\tilde{n}:0:1)$ e $\mathcal{O}$ são pontos distintos em $C_{\tilde{F}_n}(\mathbb{F}_p)$. Nos resta agora contar o número de pontos $(x:y:1) \in C_{\tilde{F}_n}(\mathbb{F}_p)$ tais que $x \neq 0, \pm\tilde{n}$. Agrupemos esses elementos em $\dfrac{p-3}{2}$ pares $\{x, -x\}$. Como $p \equiv 3 \pmod 4$ e $f(X) = X^3 - n^2X$ é uma função ímpar, então exatamente um dos elementos $f(x)$ e $f(-x) = -f(x)$ é um quadrado módulo p. Em qualquer um dos dois casos cada par fornece dois pontos em $C_{\tilde{F}_n}(\mathbb{F}_p)$ dados por $(x : \pm f(x)^{\frac{1}{2}} : 1)$ ou $(-x : \pm f(-x)^{\frac{1}{2}} : 1)$. Portanto temos $\#C_{\tilde{F}_n}(\mathbb{F}_p) = 2\dfrac{p-3}{2} + 4 = p + 1$. □

Teorema 3.3.2. *Dada a curva* $C_{F_n} : F_n(X,Y,Z) = Y^2Z - X^3 + n^2XZ^2$ *temos que* $\#C_{F_n}(\mathbb{Q})_{tor} = 4$*, ou seja, o conjunto* $C_{F_n}(\mathbb{Q})_{tor}$ *tem 4 elementos. Mais ainda, desses quatro elementos, um é o elemento neutro* $\mathcal{O}$ *e os outros três elementos têm ordem 2.*

Demonstração: O conjunto $C_{F_n}(\mathbb{Q})_{tor}$ possui, pelo menos, 4 elementos. O elemento neutro $\mathcal{O}$ e os três pontos de ordem exatamente 2, $(0:0:1), (n:0:1)$ e $(-n:0:1)$. Suponhamos que $\#C_{F_n}(\mathbb{Q})_{tor} > 4$. Logo existe $Q \in C_{F_n}(Q)$ de ordem $N > 2$. Ou seja N é ímpar ou existe $P \in C_{F_n}(Q)$ de ordem exatamente 4. No primeiro caso, seja S o subgrupo de $C_{F_n}(Q)$ gerado por Q. No segundo caso, como temos 3 pontos de ordem 2, pelo menos um destes pontos não pertence ao subgrupo gerado por P. Denotemos este ponto por R. Nesta última situação seja S o produto dos subgrupos de $C_{F_n}(Q)$ gerados por P e R. Logo $S \cong (\mathbb{Z}/4\mathbb{Z}) \times (\mathbb{Z}/4\mathbb{Z})$. Sejam $m = N$ ou 8 e $S = \{P_1, P_2, \ldots, P_m\}$.

Para cada $i, j \in \{1, \ldots, m\}$, seja $P_i = (x_i : y_i : z_i)$ tal que $x_i, y_i, z_i \in \mathbb{Z}$ e

$$P_i \times P_j = (y_iz_j - y_jz_i, x_jz_i - x_iz_j, x_iy_j - x_jy_i) \in \mathbb{R}^3$$

Se $P_i \neq P_j$ então $P_i \times P_j \neq 0$. Seja M_{ij} o máximo divisor comum das coordenadas de $P_i \times P_j$. Pela proposição 3.3.3, $\tilde{P}_i = \tilde{P}_j$ se, e somente se, p divide M_{ij}.

Seja $p > 2$ um número primo que não divide n e tal que $p > M_{ij}$. Logo $\tilde{P}_i \neq \tilde{P}_j$. Em particular, S é isomorfo via Φ_n a um subgrupo de $\tilde{C}_{F_n}(\mathbb{F}_\mathrm{p})$. Portanto, para quase todo número primo p temos que $\#\tilde{C}_{F_n}(\mathbb{F}_\mathrm{p})$ é divisível por m. Na verdade, isso permanece verdade para quase todo número primo p tal que $p \equiv 3 \pmod 4$. Pela proposição 3.3.4, temos que $\#\tilde{C}_{F_n}(\mathbb{F}_\mathrm{p}) = \mathrm{p}+1$, se $p \equiv 3 \pmod 4$. Assim, $p \equiv -1 \pmod m$ para quase todo número primo p.

O teorema das Progressões Aritméticas de Dirichlet afirma que dados dois números inteiros $r, s \neq 0$ tais que $mdc(r,s) = 1$, existem infinitos números primos da forma $rd + s$ com $d \geq 1$ inteiro. Os casos $r = 4, s = 3$ e $r = 6, s = 5$ podem ser provados da mesma forma que na demonstração do Teorema de Euclides sobre a infinidade do números de números primos. A demonstração do caso geral utiliza técnicas mais elaboradas da Teoria Analítica dos Números. Para uma demonstração completa ver capítulo 7 de [1].

Retornando à demonstração do Teorema, obtemos uma contradição com o Teorema das Progressões Aritméticas de Dirichlet tomando $r = 8$ e $s = 3$ se $m = 8$, $r = 4m$ e $s = 3$ se m é ímpar e 3 não divide m e, finalmente, $r = 12$ e $s = 7$ se m é ímpar e 3 divide m. □

3.3.2 O resultado principal

O próximo teorema faz a conexão números congruentes e a aritmética das curvas elípticas.

Teorema 3.3.3. *Um número n é congruente se, e somente se, o posto algébrico de C_{F_n} é positivo.*

Demonstração: Suponhamos que n seja um número congruente e seja (a, b) a solução da equação cúbica obtida pelo argumento que precede à Proposição 3.3.2. Neste caso temos que $a \in (\mathbb{Q}^+)^2$ com denominador par. Se (a, b) tiver ordem finita então, pelo teorema 3.3.2, temos que (a, b) é necessariamente um ponto de ordem 2. Logo sua primeira coordenada só pode ser $0, n$ ou $-n$. Claro que $0, -n \notin (\mathbb{Q}^+)^2$. Além disto para determinar se um inteiro positivo n é congruente, basta considerar sua classe módulo $(\mathbb{Q}^+)^2$. Logo supomos sempre que n é livre de quadrados. Portanto $n \notin (\mathbb{Q}^+)^2$. Logo, pelo teorema de Mordell-Weill concluimos que (a, b) tem que ser um ponto de ordem infinita de $C_{F_n}(\mathbb{Q})$. Em particular o posto algébrico de C_{F_n} é positivo.

Reciprocamente, dado um ponto $P \in C_{F_n}(\mathbb{Q})$ de ordem infinita então por 1.4.1, temos que a x-coordenada de $2P$ é

$$x_{2P} = \frac{x^4 + 4x^2n^2 + n^4}{(2y)^2}$$

e assim temos satisfeitas às condições da proposição 3.3.2 e portanto, n é um número congruente. □

Bibliografia

[1] Apostol, T., M., **Introduction to Analytic Number Theory**. Springer-Verlag, 1976.

[2] Fulton, W., **Algebraic Curves: An Introduction to Algebraic Curves**. Benjamin Cummings, 1969.

[3] Garcia, A., Lequain, Y., **Álgebra: um Curso de Introdução**. Projeto Euclides, 1988 (IMPA).

[4] Gupta, R. e Murty, M. R., **Primitive points on elliptic curves**. Compositio Math., 58 (1986), 13-44.

[5] Koblitz, N., **Introduction to Elliptic Curves and Modular Forms**. Springer-Verlag, 1984.

[6] Koblitz, N., **Elliptic curve cryptosystems**. Math. Comp., 48 (1987), 203-209.

[7] Koblitz, N., **Elliptic curve implementation of zero-knowledge blobs**. Journal of Cryptology 4 (1991), 207-213

[8] Lenstra Jr, H. W., **Factoring integers with elliptic curves**. Annals of Math. (2) 126 (1987), 649-673.

[9] Miller, V., **Use of elliptic curves in criptography**. Advances in Cryptology. Springer-Verlag, 1986, 417-426.

[10] Odlyzko, A. M., **Discrete logarithms in finite fields and their cryptographic significance**. Advances in Cryptology (Paris 1984). Springer-Verlag, 1985, 224-314

[11] Pacheco, A., **Números congruentes e curvas elíticas**. Matemática Universitária 22/23 (junho/dezembro 1997), 18-29

[12] Silverman, J., **The Arithmetic of Elliptic Curves**. Springer-Verlag, 1986

[13] Silverman, J., Tate, J., **Rational Points on Elliptic Curves**. UTM, Spriger-Verlag, 1992.

[14] Vainsencher, I., **Introdução às Curvas Algébricas**. Coleção Matemática Universitária.