USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA
FACULTADE DE MATEMÁTICAS

Traballo Fin de Grao

Curvas elípticas y aplicaciones en criptografía

Xiana Carrera Alonso

Junio, 2025

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

GRAO DE MATEMÁTICAS

Traballo Fin de Grao

Curvas elípticas y aplicaciones en criptografía

Xiana Carrera Alonso

Junio, 2025

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Trabajo propuesto

Área de Coñecemento: Álxebra
Título: Curvas elípticas e aplicacións en criptografía
Breve descrición do contido
Este traballo estudará as propiedades básicas aritméticas e xeométricas das curvas elípticas. Estudarase a súa clasificación e expoñerase a estrutura de grupo dos puntos dunha curva elíptica. Como aplicación, tratarase a aplicación dos seus resultados ao seu uso en protocolos de seguridade como as firmas dixitais ou os intercambios de chaves.
Recomendacións
Outras observacións

Índice

Resumen **viii**

Introducción **xi**

1. Preliminares sobre geometría algebraica **1**

1.1. Curvas planas afines 1

1.2. El plano proyectivo 3

1.3. Morfismos de curvas proyectivas 7

1.4. El teorema de Bézout 9

2. Curvas elípticas y la ecuación de Weierstrass **15**

2.1. Definiciones básicas sobre curvas elípticas 15

2.2. Aplicaciones entre curvas elípticas 21

2.3. El invariante j 22

3. La estructura de grupo en curvas elípticas **27**

3.1. Construcción 27

3.2. Endomorfismos 31

3.3. Puntos racionales y enteros 34

3.4. Puntos de torsión 36

4. Curvas elípticas sobre cuerpos finitos **43**

4.1. Estructura de grupo de las curvas elípticas sobre cuerpos finitos 43

4.2. El teorema de Hasse 46

5. Criptografía de curva elíptica **49**

5.1. Introducción a la criptografía de curva elíptica 49

5.2. El problema del logaritmo discreto en curvas elípticas 52

5.3. Elección de los parámetros de dominio 53

5.4. Elliptic-curve Diffie-Hellman (ECDH) 55

5.5. Sistema de criptografía asimétrica de ElGamal con curvas elípticas 55

5.6. Elliptic Curve Digital Signature Algorithm (ECDSA) 56

I. El invariante j a través de transformaciones proyectivas **59**

II. Ataques al problema del logaritmo discreto **63**

II.1. Algoritmo de Pohlig-Hellman 63

II.2. Algoritmo ρ de Pollard 65

Bibliografía **69**

Resumen

El objetivo de este trabajo es ofrecer un estudio exhaustivo sobre las curvas elípticas, un caso particular de curvas algebraicas que ha ocupado un lugar destacado en diversas ramas de las matemáticas, como la geometría algebraica y la teoría de números, y que ha encontrado importantes aplicaciones en la criptografía moderna. Con el fin de ofrecer un análisis detallado tanto de los aspectos teóricos como de sus aplicaciones, el trabajo comienza introduciendo conceptos y resultados de geometría algebraica que constituyen el marco fundamental para el desarrollo posterior. A continuación, se presenta tanto la definición formal de curva elíptica como su clasificación en función del invariante j. Tras abordar una de sus más importantes características, su estructura de grupo, se examinan las propiedades teóricas de las curvas elípticas en cuerpos finitos, que son el objeto de interés para la parte final, en la que se explora su uso práctico en la criptografía.

Abstract

The aim of this work is to provide a thorough study of elliptic curves, a particular case of algebraic curves that has occupied a prominent position in several branches of mathematics, such as algebraic geometry and number theory, and that has found significant applications in modern cryptography. In order to provide a detailed analysis on both the theoretical aspects and their applications, this bachelor thesis begins by introducing key concepts and results from algebraic geometry that serve as the fundamental framework for the subsequent development. Next, the formal definition of an elliptic curve is presented, as well as its classification based on the j-invariant. After discussing one of its most notable properties, that being its group structure, the focus shifts to the theoretical properties of elliptic curves over finite fields, which are the main object of interest in the final part, where their practical application in cryptography is explored.

Introducción

La primera aparición de las curvas elípticas tiene lugar en los trabajos de Diofanto de Alejandría, entre los siglos II y III a. C., pero su nombre no fue acuñado hasta el siglo XIX. El comienzo de la historia de esta denominación se remonta al siglo XVIII, cuando Giulio Fagnano y Euler inician un intenso estudio sobre las longitudes de arcos de elipses, que vienen dadas por integrales para las que, tras un cambio de variable elemental, el integrando se puede expresar como una función racional que involucra la raíz de un polinomio de grado 3 o 4. En la actualidad, sabemos que estas integrales, que recibieron el nombre de *integrales elípticas*, en general no tienen solución en términos de funciones elementales [32].

Cuando se considera un círculo, la integral resultante involucra en su lugar un polinomio cuadrático, y su solución se puede expresar en términos de la función arcoseno. La inversa de la integral viene dada por la función seno, cuyas propiedades son bien notas. Siguiendo esta misma idea, Gauss propone en su diario (no más allá de 1796) invertir las integrales elípticas, cuyo resultado son funciones que se denominan *funciones elípticas*. Abel es el primero en publicar un artículo sobre estas, y tanto él como Gauss y Jacobi desarrollan la teoría incipiente [16].

Estas funciones (que, bajo una perspectiva moderna, son funciones complejas, meromorfas y doblemente periódicas) parametrizan las llamadas *curvas elípticas*. No es hasta 1901 que Poincaré unifica y generaliza estos trabajos mediante la introducción de las *curvas algebraicas*, y en la actualidad el estudio de las curvas elípticas se puede enmarcar dentro del campo de la geometría algebraica. Sin embargo, por sus ricas propiedades aritméticas, son también un importante objeto de estudio en la teoría de números, siendo relevantes en numerosos trabajos de investigación actuales [33]. Notablemente, son uno de los objetos centrales de la conjetura de Birch y Swinnerton-Dyer, que es uno de los siete problemas del milenio propuestos por el Instituto Clay de Matemáticas en el año 2000. Cabe destacar también la relevancia de las curvas elípticas en la prueba de Andrew Wiles del último teorema de Fermat, y en particular en la demostración del teorema (previamente conjetura) de Taniyama-Shimura, que establece una relación entre las curvas elípticas y las formas modulares y constituye uno de los pilares fundamentales en la prueba de Wiles [9].

Desde finales del siglo XX, las curvas elípticas han adquirido además un protagonismo crecien-

te en la criptografía. Los trabajos independientes de Neal Koblitz y Victor Miller en 1985 [23, 27] sentaron las bases de la criptografía basada en curvas elípticas, que se ha consolidado como una alternativa atractiva a sistemas clásicos como RSA por requerir claves más cortas para alcanzar el mismo nivel de seguridad, con importantes ventajas en términos de eficiencia y velocidad.

Con el objetivo de ofrecer un estudio que recoja las propiedades más importantes de las curvas elípticas, en la exposición se adopta un enfoque moderno que requiere de conocimientos previos de geometría algebraica, a los cuales se dedica el Capítulo 1. Este presenta conceptos como la noción de curva, los espacios proyectivos, los morfismos de curvas proyectivas y el teorema de Bézout.

En el Capítulo 2 se da una definición formal de curva elíptica, estableciendo su relación con definiciones alternativas que se pueden encontrar en la literatura. Además, se demuestra que, descontando ciertas excepciones, las curvas elípticas quedan especificadas por una forma simplificada de la conocida como ecuación de Weierstrass. También se comentan brevemente las isogenias (morfismos particulares entre curvas elípticas) y se analiza el invariante j, que permite clasificar las curvas elípticas salvo isomorfismo.

El Capítulo 3 introduce la estructura de grupo en las curvas elípticas, dando tanto una interpretación geométrica como una demostración formal de su existencia y de sus principales propiedades. A partir de esta estructura, se profundiza sobre las isogenias, hablando ahora de endomorfismos. Asimismo, se presentan distintos teoremas que caracterizan el grupo de una curva elíptica, prestando especial atención a sus puntos racionales y enteros, lo que permite establecer una conexión con los trabajos aritméticos de Diofanto de Alejandría. Finalmente, se realiza un estudio los puntos de torsión y se prueba el teorema de Lutz-Nagell, que ilustra el uso de herramientas de teoría de números en el estudio de las curvas elípticas.

El Capítulo 4 se dedica a las curvas elípticas sobre cuerpos finitos, que son particularmente interesantes por sus aplicaciones criptográficas. Se describen las características específicas de su estructura de grupo y, empleando los resultados desarrollados previamente sobre endomorfismos, se prueba el teorema de Hasse, que da una cota para el número de puntos de una curva elíptica definida sobre un cuerpo finito.

Por último, el Capítulo 5 se centra en las aplicaciones criptográficas. Tras introducir brevemente conceptos fundamentales de criptografía, se explica el problema del logaritmo discreto en curvas elípticas, que es la base matemática de sus sistemas criptográficos. A continuación, se indican las principales medidas que se deben tomar al elegir parámetros de este problema para garantizar la seguridad de los algoritmos que se basan en él. Finalmente, se presentan tres ejemplos de protocolos criptográficos que emplean curvas elípticas con diferentes objetivos en el contexto de las comunicaciones digitales: el protocolo de intercambio de claves Diffie-Hellman (ECDH), el protocolo de criptografía asimétrica de ElGamal y un algoritmo de firma digital (ECDSA).

Capítulo 1

Preliminares sobre geometría algebraica

Este capítulo presenta conceptos y resultados de geometría algebraica que servirán a modo de punto de partida para el posterior estudio de las curvas elípticas. Tras definir el concepto de curvas afines, se introducen las curvas proyectivas, analizadas como subconjuntos de los espacios proyectivos. A continuación, se realiza un estudio de la intersección de curvas de modo algebraico, que culmina exponiendo el clásico teorema de Bézout.

Las principales referencias son los libros de Milne [7], Kunz [6], Fulton [1] y Silverman [8], así como las notas de Gathmann [3] y de Sutherland [36].

1.1. Curvas planas afines

Las curvas suponen un punto de encuentro entre el mundo algebraico y el geométrico. Representan conjuntos de soluciones de ecuaciones polinómicas, por lo que dan pie a un desarrollo doble: como objetos algebraicos, a través de las características de los polinomios que las definen, y como objetos geométricos, a través de las propiedades geométricas de los conjuntos que representan dentro de un determinado espacio.

A lo largo de este trabajo, fijaremos la siguiente notación: K denotará un cuerpo, y $\overline{K}$, una clausura algebraica de K. $\mathbb{A}^n(K) := K^n$ será el *espacio afín* de dimensión $n \in \mathbb{N}$ sobre K, aunque, si no hay ambigüedad, escribiremos simplemente $\mathbb{A}^n$. En particular, $\mathbb{A}^1 = K$ se conoce como *recta afín* y $\mathbb{A}^2 = K^2$, como *plano afín*.

Definición 1.1. Un subconjunto $C \subset \mathbb{A}^2$ es una *curva (algebraica plana afín)* si existe un polinomio

$f \in K[x,y]$ no constante y sin factores repetidos en $\overline{K}[x,y]$ tal que

$$C = \{(x,y) \in \mathbb{A}^2 \mid f(x,y) = 0\}.$$

En tal caso, escribiremos $C : f = 0$. Emplearemos también la notación C_f.

Nótese que, en la anterior definición, existen varias posibilidades para la elección de f, pues $C_f = C_{\lambda f}$ para todo $\lambda \in K^\times$. Por tanto, la curva C corresponde a una clase de la relación de equivalencia de polinomios no constantes en $K[x,y]$ y sin factores repetidos en $\overline{K}[x,y]$ dada por $f \sim g \iff f = \lambda g$ para algún $\lambda \in K^\times$.

Introducimos a continuación la noción de irreducibilidad de una curva:

Definición 1.2. Se dice que una curva C_f es *irreducible* si f es irreducible en $K[x,y]$.

Dada una curva C_f, podemos escribir $f = f_1 f_2 \dots f_r$ con $f_i \in K[x,y]$ polinomios irreducibles distintos, y obtener que $C_f = C_{f_1} \cup \dots \cup C_{f_r}$, donde cada C_{f_i} es una curva irreducible. Las curvas C_{f_i} se denominan *componentes irreducibles* de C_f.

Emplearemos también el concepto de grado de una curva, definido como sigue:

Definición 1.3. Sea C_f una curva. El *grado* de C_f es el grado de f como polinomio en $K[x,y]$. Las curvas de grados 1, 2, 3 y 4 reciben el nombre de *rectas*, *cónicas*, *cúbicas* y *cuárticas*, respectivamente.

Observación 1.4. Tanto el concepto de irreducibilidad como el de grado de una curva son independientes de la elección del representante de la clase de equivalencia de f y, por tanto, están bien definidos. Lo son también las nociones que introduciremos a continuación, por lo que no volveremos a hacer mención de este punto.

Definición 1.5. Sea $K_0 \subset K$ un subanillo de K. Si $C : f = 0$ es una curva con $f \in K_0[x,y]$ un polinomio no constante y sin factores repetidos en $\overline{K_0}[x,y]$, se dice que C está *definida sobre* K_0.

El conjunto $C(K_0) := C \cap K_0^2$ se denomina *conjunto de puntos K_0-racionales de C*.

Definición 1.6. Consideremos una curva $C_f \subset \mathbb{A}^2$ y un punto $P \in C_f$. Diremos que P es un *punto singular* de C_f si se cumple que

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

En caso contrario, diremos que P es un *punto regular* de C_f.

Si todos los puntos de C_f son regulares, se dice que C_f es una curva *no singular* o *lisa*. De lo contrario, se dice que es *singular*.

Definición 1.7. Sea $C_f \subset \mathbb{A}^2$ una curva, y sea $P = (x_0, y_0) \in C_f$ un punto regular. La *recta tangente* a C_f en P es la dada por la ecuación

$$(x - x_0)\frac{\partial f}{\partial x}(P) + (y - y_0)\frac{\partial f}{\partial y}(P) = 0.$$

Observación 1.8. Un punto en la intersección de dos componentes irreducibles de una curva es necesariamente singular, pues si C_f es una curva con componentes irreducibles C_{f_1} y C_{f_2}, de modo que $f = f_1 f_2$, y $P \in \mathbb{A}^2$ es tal que $f_1(P) = f_2(P) = 0$, entonces $\frac{\partial f}{\partial x}(P) = \frac{\partial f_1}{\partial x}(P) f_2(P) + f_1(P)\frac{\partial f_2}{\partial x}(P) = 0$, y análogamente $\frac{\partial f}{\partial y}(P) = \frac{\partial f_1}{\partial y}(P) f_2(P) + f_1(P)\frac{\partial f_2}{\partial y}(P) = 0$.

1.2. El plano proyectivo

En la presente sección se describen los espacios proyectivos y, en particular, el que nos será de mayor interés, el plano proyectivo. Este será el espacio en el que definiremos las curvas elípticas, pues algunas de sus propiedades más notables, como el grupo asociado a las mismas, dependen de conceptos proyectivos.

Definición 1.9. El *espacio proyectivo n-dimensional* sobre un cuerpo K, denotado por $\mathbb{P}^n(K)$ o $\mathbb{P}^n$ si no hay ambigüedad, es el conjunto de subespacios de dimensión uno de $\mathbb{A}^{n+1}$. Equivalentemente, es el conjunto de $(n+1)$-tuplas no nulas $(x_1, \dots, x_n, x_{n+1}) \in \mathbb{A}^{n+1}$ módulo la relación de equivalencia dada por $(x_1, \dots, x_n, x_{n+1}) \sim (y_1, \dots, y_n, y_{n+1})$ si $\exists \lambda \in K \setminus \{0\}$ tal que $(x_1, \dots, x_n, x_{n+1}) = \lambda(y_1, \dots, y_n, y_{n+1})$.

Denotaremos los elementos (o puntos) de $\mathbb{P}^n$ como $(x_1 : \dots : x_n : x_{n+1})$, donde $x_1, \dots, x_n, x_{n+1} \in K$, conocidas como *coordenadas homogéneas*, son no todas nulas.

Observación 1.10. Existe una inyección

$$\begin{aligned} \mathbb{A}^n &\hookrightarrow \mathbb{P}^n \\ (x_1, \dots, x_n) &\longmapsto (x_1 : \dots : x_n : 1) \end{aligned}$$

cuya imagen, $U_{n+1} := \{(x_1 : \dots : x_n : x_{n+1}) \in \mathbb{P}^n \mid x_{n+1} \neq 0\}$ puede ser identificada con $\mathbb{A}^n$ (véase la Figura 1.1). Análogamente, definimos $U_i := \{(x_1 : \dots : x_n : x_{n+1}) \in \mathbb{P}^n \mid x_i \neq 0\}$ para $i = 1, \dots, n, n+1$. Estos conjuntos recubren $\mathbb{P}^n$ si $K = \mathbb{R}$ o $K = \mathbb{C}$ y lo dotan de una estructura de variedad topológica de dimensión n, sirviendo a modo de cartas coordenadas junto a los respectivos homeomorfismos

$$\begin{aligned} U_i &\xrightarrow{\cong} \mathbb{A}^n \\ (x_1 : \dots : x_n : x_{n+1}) &\longmapsto \left(\frac{x_1}{x_i}, \dots, \frac{x_{i-1}}{x_i}, \frac{x_{i+1}}{x_i}, \dots, \frac{x_n}{x_i}, \frac{x_{n+1}}{x_i}\right). \end{aligned}$$

Los conjuntos $\mathbb{P}^n \setminus U_i = \{(x_1 : \dots : x_n : x_{n+1}) \in \mathbb{P}^n \mid x_i = 0\}$ se identifican con un espacio proyectivo de dimensión $n-1$, y podemos interpretarlos como un hiperplano de puntos del infinito. Así, es posible identificar $\mathbb{P}^n$ con la unión disjunta $\mathbb{A}^n \sqcup \mathbb{P}^{n-1}$. En particular, para el caso de la recta proyectiva $\mathbb{P}^1$, tenemos que $\mathbb{P}^1 = \mathbb{A}^1 \sqcup \mathbb{P}^0$, donde $\mathbb{P}^0 = \{(0 : 1)\}$ es un conjunto de un solo punto:

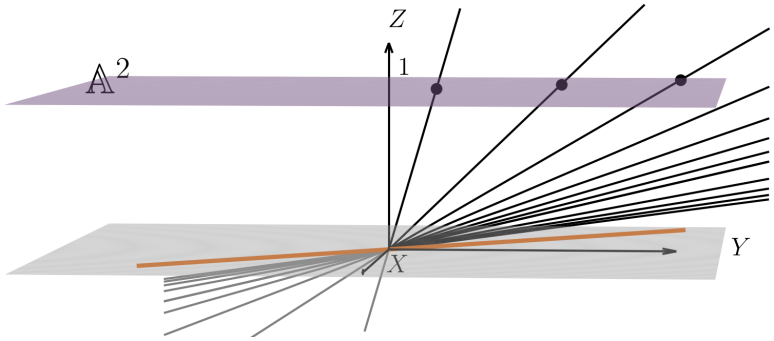


Figura 1.1: Correspondencia entre un punto $(x, y) \in \mathbb{A}^2$, la recta que pasa por $(x, y, 1) \in K^3$ y por $(0, 0, 0) \in K^3$, y el punto $(x : y : 1) \in \mathbb{P}^2$. Obsérvese que los puntos del infinito no tienen esta correspondencia. En su lugar, pueden ser vistos como límites de rectas que pasan por el origen y por una secuencia de puntos en $\mathbb{A}^2$.

el punto del infinito ∞. Si tomamos $K = \mathbb{R}$, podemos visualizar $\mathbb{P}^1$ como la compactificación de Alexandroff de la recta real. A su vez, $\mathbb{P}^2 = \mathbb{A}^2 \sqcup \mathbb{P}^1$ puede verse como el plano afín con un punto del infinito añadido para cada dirección en el plano.

En general, el valor de un polinomio $f \in K[X_1, \dots, X_n, X_{n+1}]$ en un punto $P = (x_1 : \dots : x_n : x_{n+1}) \in \mathbb{P}^n$ no está bien definido, pues depende del representante de P. Sin embargo, si f es *homogéneo* (es decir, si todos los monomios de f tienen el mismo grado), se tiene que

$$f(\lambda X_1, \dots, \lambda X_n, \lambda X_{n+1}) = \lambda^d f(X_1, \dots, X_n, X_{n+1}),$$

donde $d = \deg f$ y $\lambda \in K$. En tal caso, la condición $f(P) = 0$ sí está bien definida, lo que nos lleva al concepto de *curva proyectiva*. Para su definición nos restringiremos al caso que nos será de interés, el plano proyectivo $\mathbb{P}^2$, por lo que hablaremos de curvas proyectivas *planas*.

Definición 1.11. Se dice que un subconjunto $C \subset \mathbb{P}^2$ es una *curva (algebraica plana) proyectiva* si existe un polinomio homogéneo $F \in K[X, Y, Z]$ sin factores repetidos en $\overline{K}[X, Y, Z]$ tal que

$$C = \{(x : y : z) \in \mathbb{P}^2 \mid F(x, y, z) = 0\}.$$

Utilizaremos la notación $C : F = 0$ o sencillamente C_F.

Análogamente al caso afín, diremos que C_F es *irreducible* si F es irreducible en $K[X, Y, Z]$. Si $F = F_1 F_2 \dots F_r$ con $F_i \in K[X, Y, Z]$ polinomios homogéneos irreducibles distintos, entonces C_F es la unión de las curvas proyectivas C_{F_i}, que son sus *componentes irreducibles*.

El *grado* de C_F es el grado de F como polinomio en $K[X, Y, Z]$.

Como en el caso afín, hacemos notar que una curva proyectiva se corresponde con una clase de la relación de equivalencia de polinomios en $K[X, Y, Z]$ sin factores repetidos en $\overline{K}[X, Y, Z]$ dada por

$F \sim G \iff F = \lambda G$ para algún $\lambda \in K^\times$. Las nociones de irreducibilidad, componentes irreducibles y grado, así como todas las que introduciremos a continuación, son consistentes con este hecho.

Las curvas proyectivas de grado uno en $\mathbb{P}^2$ reciben el nombre de *rectas proyectivas*. Pueden ser descritas como el conjunto de ceros de polinomios homogéneos $F \in K[X,Y,Z]$ de grado uno, es decir, conjuntos de soluciones de ecuaciones del tipo

$$aX + bY + cZ = 0, \quad (a,b,c) \in K^3 \setminus \{(0,0,0)\}.$$

Una consecuencia inmediata es que dos rectas proyectivas distintas se cortan en un único punto (lo cual no es cierto en el caso afín, donde puede haber rectas paralelas), debido a que en el plano proyectivo formarían un sistema homogéneo de dos ecuaciones lineales en tres incógnitas. Si las rectas son distintas, el rango de la matriz asociada es dos, por lo que existe una única solución salvo multiplicación por un escalar no nulo.

Las curvas proyectivas de grados dos, tres y cuatro se conocen como *cónicas*, *cúbicas* y *cuárticas* proyectivas, respectivamente.

Análogamente al caso afín, tenemos la siguiente definición:

Definición 1.12. Sea $K_0 \subset K$ un subanillo de K. Si $C : F = 0$ es una curva proyectiva con $F \in K_0[X,Y,Z]$ un cierto polinomio homogéneo no constante y sin factores repetidos en $\overline{K_0}[X,Y,Z]$, se dice que C está *definida sobre* K_0.

El conjunto de puntos $P \in C$ que se pueden expresar como $P = (x : y : z)$, con $x, y, z \in K_0$, se denomina *conjunto de puntos K_0-racionales de C*.

Dado un polinomio $f \in K[X_1, \dots, X_n]$ de grado d, podemos construir un correspondiente polinomio homogéneo $\widehat{f} \in K[X_1, \dots, X_n, X_{n+1}]$ de grado d. Esto nos será de utilidad a la hora de relacionar curvas proyectivas y afines, que a su vez nos permitirá emplear numerosos resultados de la geometría afín para el estudio de las curvas proyectivas.

Definición 1.13. Sea $f \in K[X_1, \dots, X_n]$ un polinomio de grado d. Se denomina *homogeneización* de f (con respecto a X_{n+1}) al polinomio homogéneo de grado d dado por

$$\widehat{f}(X_1, \dots, X_n, X_{n+1}) = X_{n+1}^d f\left(\frac{X_1}{X_{n+1}}, \dots, \frac{X_n}{X_{n+1}}\right) \in K[X_1, \dots, X_n, X_{n+1}].$$

En $\mathbb{P}^2$, es común considerar la homogeneización de un polinomio $f \in K[X,Y]$ de grado d con respecto a la variable Z.

La recta proyectiva $Z = 0$ en $\mathbb{P}^2$ recibe el nombre de *recta del infinito*, y los puntos que satisfacen tal ecuación se denominan *puntos del infinito*. Obsérvese que la homogeneización $\widehat{f} \in K[X,Y,Z]$ de

un polinomio $f \in K[X,Y]$ de grado d con respecto a Z contiene un monomio en el que no aparece la variable Z (el término de mayor grado de f), por lo que ecuación de la recta del infinito no divide a $\widehat{f}$.

Definición 1.14. Sea $F \in K[X_1,\dots,X_n,X_{n+1}]$ un polinomio homogéneo de grado d. Se denomina *deshomogeneización* de F (con respecto a X_{n+1}) al polinomio dado por

$$\widehat{F}(X_1,\dots,X_n) = F(X_1,\dots,X_n,1) \in K[X_1,\dots,X_n].$$

Observación 1.15. Con la notación de la anterior definición, si F no es divisible por X_{n+1}, entonces $\deg \widehat{F} = \deg F$. Por tanto, existe una correspondencia biunívoca entre los polinomios de grado d en $K[X_1,\dots,X_n]$ y los polinomios homogéneos de grado d en $K[X_1,\dots,X_n,X_{n+1}]$ que no son divisibles por X_{n+1}.

Definición 1.16. Sea $f \in K[x,y]$ un polinomio no constante sin factores repetidos en $\overline{K}[x,y]$, y sea $C : f = 0$ una curva afín. La *clausura proyectiva* de C es el conjunto de ceros de la homogeneización de f, esto es, el conjunto

$$\widehat{C} := \left\{ P \in \mathbb{P}^2 : \widehat{f}(P) = 0 \right\} \subset \mathbb{P}^2.$$

Razonando como en la observación 1.10, el plano proyectivo $\mathbb{P}^2$ puede ser recubierto por tres planos afines, para los cuales emplearemos la siguiente notación:

Notación 1.17. Denotaremos por V_1, V_2 y V_3 a los planos afines $V_1 := \{(1 : y : z) \mid y,z \in K\}$, $V_2 := \{(x : 1 : z) \mid x,z \in K\}$ y $V_3 := \{(x : y : 1) \mid x,y \in K\}$.

Proposición 1.18. *Identificando $\mathbb{A}^2$ con un conjunto $V_i \subset \mathbb{P}^2$, $i = 1,2,3$, se cumplen las siguientes propiedades:*

(i) Sea C una curva afín. Su clausura proyectiva $\widehat{C}$ es una curva proyectiva y $\widehat{C} \cap \mathbb{A}^2 = C$.

(ii) Sea C una curva proyectiva. Entonces $C \cap \mathbb{A}^2$ es una curva afín y cumple que o bien $C \cap \mathbb{A}^2 = \emptyset$ o bien $\widehat{C \cap \mathbb{A}^2} = C$.

(iii) Si C es una curva afín (respectivamente, proyectiva) definida sobre K_0, entonces $\widehat{C}$ (respectivamente, $C \cap \mathbb{A}^2$) también está definida sobre K_0.

Demostración. Se remite al lector a [18, Corollary I.2.3] para las demostraciones de (i) e (ii). La propiedad (iii) es inmediata de la definición de clausura proyectiva. □

Ejemplo 1.19. Consideremos una curva afín de la forma $C : y^2 + a_1xy + a_3y - x^3 - a_2x^2 - a_4x - a_6 = 0$, con $a_1,\dots,a_6 \in K$. Es inmediato ver que su clausura proyectiva es la curva proyectiva $\widehat{C} : Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3$, que contiene a C y el punto del infinito $(0 : 1 : 0)$. Trataremos este ejemplo con más detalle en el capítulo 2.

De forma análoga al caso afín, en el caso proyectivo definimos los puntos singulares y las rectas tangentes a puntos regulares como sigue:

Definición 1.20. Consideremos una curva proyectiva $C_F \subset \mathbb{P}^2$ y un punto $P = (x : y : z) \in C_F$. Diremos que P es un *punto singular* de C_F si se cumple que

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

En caso contrario, diremos que P es un *punto regular* o *no singular* de C_F.

Si todos los puntos de C_F son regulares, se dice que C_F es una curva *no singular* o *lisa*. De lo contrario, se dice que es *singular*.

Definición 1.21. Sea $C_F \subset \mathbb{P}^2$ una curva proyectiva, y sea $P = (x : y : z) \in C_F$ un punto regular. La *recta tangente proyectiva* a C_F en P es la dada por la ecuación

$$(X - x)\frac{\partial F}{\partial X}(P) + (Y - y)\frac{\partial F}{\partial Y}(P) + (Z - z)\frac{\partial F}{\partial Z}(P) = 0.$$

Es interesante notar la relación entre las rectas tangentes del caso proyectivo y las del caso afín. Sean C una curva proyectiva y $V_i \subset \mathbb{P}^2$, $i = 1, 2, 3$. Por la proposición 1.18, sabemos que $C \cap V_i$ es una curva afín. Si $P \in C \cap V_i$ es un punto regular de C con recta tangente proyectiva L, entonces se tiene que $L \cap V_i$ es la recta tangente afín a C en P.

1.3. Morfismos de curvas proyectivas

A continuación, definiremos los *morfismos de curvas proyectivas*, que son las aplicaciones que preservan la estructura de curvas proyectivas. Esto nos servirá como base para estudiar, en posteriores capítulos, el caso particular de morfismos entre curvas elípticas (llamados *isogenias*).

A lo largo de esta sección supondremos que K es algebraicamente cerrado.

Definición 1.22. Sea $C : F = 0$ una curva proyectiva con $F \in K[X, Y, Z]$ un polinomio irreducible. Definimos el *cuerpo de funciones racionales* de C como:

$$K(C) = \left\{ \frac{G}{H} : G, H \in K[X, Y, Z] \text{ homogéneos del mismo grado, } H \notin (F) \right\} / \sim,$$

con $G_1/H_1 \sim G_2/H_2$ si y solo si $G_1H_2 - G_2H_1 \in (F)$, siendo (F) el ideal generado por F en $K[X, Y, Z]$.

Observación 1.23. $K(C)$ es un cuerpo con las operaciones de suma y multiplicación de funciones racionales, y contiene K como subcuerpo. De hecho, se puede demostrar que $K(C)/K$ es una extensión de cuerpos trascendental con grado de trascendencia uno.

Definición 1.24. Sean C una curva proyectiva y $\alpha \in K(C)$. Se dice que α está *definida* o que es *regular* en un punto $P \in C$ si α puede ser representado como G/H para ciertos $G, H \in K[X, Y, Z]$ homogéneos del mismo grado y con $H(P) \neq 0$.

Si esto ocurre, escribimos $\alpha(P) := G(P)/H(P)$.

Observación 1.25. El valor $\alpha(P)$ está bien definido, por los siguientes motivos:

1. Dados $G, H \in K[X, Y, Z]$ dos polinomios homogéneos del mismo grado d, con $H(P) \neq 0$, entonces $G(P)/H(P)$ no depende de la elección del representante de la clase de equivalencia de un punto $P \in \mathbb{P}^2$, debido a que
$$\frac{G(\lambda P)}{H(\lambda P)} = \frac{\lambda^d G(P)}{\lambda^d H(P)} = \frac{G(P)}{H(P)}$$
para todo $\lambda \in K^\times$.
2. Si $G_1/H_1, G_2/H_2 \in K(C)$ son tales que $G_1/H_1 \sim G_2/H_2$, con $H_1(P), H_2(P) \neq 0$, entonces $G_1(P)H_2(P) - G_2(P)H_1(P)$ es un múltiplo de $F(P) = 0$, por lo que $(G_1/H_1)(P) = (G_2/H_2)(P)$.

Nótese que aunque $\alpha = G_1/H_1 \in K(C)$ con $H_1(P) = 0$, puede ocurrir que exista $G_2/H_2 \in K(C)$ tal que $G_1/H_1 \sim G_2/H_2$ y con $H_2(P) \neq 0$. No obstante, no siempre hay un modo directo de encontrar tal G_2/H_2, ya que, en general, no hay una forma de simplificar G_1/H_1 a su mínima expresión, puesto que el anillo $K[X, Y, Z]/(F)$ puede no ser un dominio de factorización única.

Estamos ahora en condiciones de definir aplicaciones racionales entre curvas proyectivas:

Definición 1.26. Sean C_1 y C_2 curvas proyectivas. Una *aplicación racional* de C_1 a C_2 es una terna $\phi = (\phi_X : \phi_Y : \phi_Z) \in \mathbb{P}^2(K(C_1))$ tal que $(\phi_X(P) : \phi_Y(P) : \phi_Z(P)) \in C_2$ para todo $P \in C_1$ en el que $\phi_X(P)$, $\phi_Y(P)$ y $\phi_Z(P)$ estén definidos y no sean todos nulos.

Decimos que ϕ está *definida* o que es *regular* en $P \in C_1$ si existe $\lambda \in K(C)^\times$ tal que $\lambda\phi_X$, $\lambda\phi_Y$ y $\lambda\phi_Z$ están definidos en P y no son todos nulos en P. Si tal λ existe, definimos $\phi(P) = ((\lambda\phi_X)(P) : (\lambda\phi_Y)(P) : (\lambda\phi_Z)(P))$.

Observación 1.27. Nótese que, en la definición de aplicación racional regular en un punto, puede ser necesario tomar un $\lambda \in K(C)^\times$ diferente para puntos $P \in C_1$ diferentes.

Definición 1.28. Sean C_1 y C_2 curvas proyectivas. Una aplicación racional de C_1 a C_2 que está definida en todo punto de C_1 se denomina *morfismo de curvas*.

En el caso de curvas elípticas, es innecesario distinguir entre aplicaciones racionales y morfismos, pues todas las aplicaciones racionales están definidas en todo punto. De forma general, se tiene:

Teorema 1.29. *Sea C_1 una curva proyectiva no singular. Entonces, toda aplicación racional de C_1 a una curva proyectiva C_2 es un morfismo.*

Demostración. Véase [8, Proposition II.2.1]; aunque es sencilla, hace uso de conceptos de álgebra conmutativa que no hemos definido. □

1.4. El teorema de Bézout

Aunque el plano proyectivo $\mathbb{P}^2$ se construye con el objetivo de que dos rectas proyectivas distintas se corten en un punto [9, 1], el famoso teorema de Bézout nos proporciona una generalización de este resultado, estableciendo una relación exacta entre los grados de dos curvas proyectivas de grado arbitrario y el número de puntos de intersección de las mismas, contados con multiplicidad. Dedicaremos esta sección a formalizar el concepto de la multiplicidad de un punto de intersección, primero en el caso afín y a continuación en el proyectivo, y a enunciar el teorema de Bézout, que nos garantiza que la recta proyectiva que une dos puntos de una curva elíptica la interseca siempre en un tercer punto, y en ningún otro más. Esto será fundamental para dotar a las curvas elípticas de una estructura de grupo en el capítulo 3,

Definición 1.30. (i) Sea $P \in \mathbb{A}^2$. Definimos el *anillo local* de P en $\mathbb{A}^2$ como el conjunto

$$\mathcal{O}_P(\mathbb{A}^2) := \left\{ \frac{f}{g} \,:\, f, g \in K[x,y],\ g(P) \neq 0 \right\} \subset K(x,y),$$

donde $K(x,y)$ es el cuerpo de fracciones de $K[x,y]$.

(ii) Definimos el *homomorfismo de evaluación* en P como la aplicación

$$\mathcal{O}_P(\mathbb{A}^2) \longrightarrow K, \quad \frac{f}{g} \longmapsto \frac{f(P)}{g(P)},$$

y denotamos su núcleo por $\mathfrak{m}_P(\mathbb{A}^2) := \{ f/g \in \mathcal{O}_P(\mathbb{A}^2) \,:\, f(P) = 0 \}$.

Observación 1.31. (a) Es sencillo comprobar que, en efecto, $\mathcal{O}_P(\mathbb{A}^2)$ es un subanillo de $K(x,y)$.

(b) La imagen de un elemento $\phi \in \mathcal{O}_P(\mathbb{A}^2)$ por el homomorfismo de evaluación en P está bien definida, pues si $\phi = f/g = f'/g'$, con $g(P)\,g'(P) \neq 0$, entonces $fg' - f'g = 0$, de modo que $f(P)\,g'(P) - f'(P)\,g(P) = (fg' - f'g)(P) = 0$, lo que implica que $f(P)/g(P) = f'(P)/g'(P)$.

(c) $\mathcal{O}_P(\mathbb{A}^2)$ es un anillo "local" también en el sentido algebraico, i.e., tiene un único ideal maximal: como $\mathfrak{m}_P(\mathbb{A}^2) = \{ \phi \in \mathcal{O}_P(\mathbb{A}^2) \,:\, \phi \text{ no es una unidad de } \mathcal{O}_P(\mathbb{A}^2) \}$, tenemos que cualquier ideal propio de $\mathcal{O}_P(\mathbb{A}^2)$ está contenido en $\mathfrak{m}_P(\mathbb{A}^2)$, y por tanto este es el único ideal maximal.

(d) Las propiedades "locales" de la curva que dependen solo en un entorno de P se reflejan en $\mathcal{O}_P(\mathbb{A}^2)$. [1]

[1] $\mathcal{O}_P(\mathbb{A}^2)$ es el anillo de gérmenes (de funciones) en P del haz que asigna a cada abierto $U \subset \mathbb{A}^2$, con la topología de Zariski, el conjunto de funciones regulares en U.

Definición 1.32. Sean $\mathbb{A}^2$ el plano afín sobre un cuerpo K, C_f y C_g dos curvas planas afines y $P \in \mathbb{A}^2$. Definimos la *multiplicidad de la intersección* de C_f y C_g en el punto P como

$$\mu_P(C_f, C_g) := \dim_K \left(\mathcal{O}_P(\mathbb{A}^2)/(f,g)\right) \in \mathbb{N} \cup \{\infty\},$$

donde $\dim_K$ denota la dimensión como espacio vectorial sobre K. Por simplicidad, emplearemos también la notación $\mu_P(f,g)$ para denotar $\mu_P(C_f, C_g)$.

Esta definición plantea una serie de cuestiones inmediatas, como cuándo la multiplicidad de la intersección es finita o qué ocurre cuando C_f o C_g no contienen a P. Recogemos a continuación algunas propiedades de la multiplicidad de la intersección.

Proposición 1.33. *Con la notación de la definición precedente, se cumplen las siguientes propiedades:*

(i) $\mu_P(f,g) = \mu_P(g,f)$.

(ii) $\mu_P(f,g) = 0 \iff P \notin C_f \cap C_g$.

(iii) $\mu_P(f,g) = 1 \iff \mathfrak{m}_P(\mathbb{A}^2) = (f,g)$.

(iv) $\mu_P(f,g) < \infty \iff$ *C_f y C_g no tienen componentes irreducibles comunes que contengan a P.*

Demostración. (i) Sigue de que $(f,g) = (g,f)$.

(ii) ($\Longleftarrow$) Supongamos que $P \notin C_f$, es decir, $f(P) \neq 0$. Entonces $f \notin \mathfrak{m}_P(\mathbb{A}^2)$. Equivalentemente, f es una unidad en $\mathcal{O}_P(\mathbb{A}^2)$, de modo que $(f,g) = \mathcal{O}_P(\mathbb{A}^2)$ y, por tanto, $\mu_P(f,g) = 0$. El caso $P \notin C_g$ es análogo.

($\Longrightarrow$) Supongamos que $P \in C_f \cap C_g$. Entonces $f(P) = g(P) = 0$ y el homomorfismo de evaluación pasa al cociente como una aplicación K-lineal $\mathcal{O}_P(\mathbb{A}^2)/(f,g) \longrightarrow K$, que es sobreyectiva. Por tanto, $\mu_P(f,g) > 0$.

(iii) Por (ii), podemos asumir que $P \in C_f \cap C_g$. Entonces, $\mu_P(f,g) = 1$ si y solo si la aplicación $\mathcal{O}_P(\mathbb{A}^2)/(f,g) \longrightarrow K$ inducida por el homomorfismo de evaluación en P es un isomorfismo de K-espacios vectoriales, lo que equivale a que (f,g) sea el núcleo del homomorfismo de evaluación en P, esto es, $\mathfrak{m}_P(\mathbb{A}^2) = (f,g)$.

(iv) Puede consultarse en [6, Chapter 5]. □

Ejemplo 1.34. Por definición, $\mathfrak{m}_{(0,0)}(\mathbb{A}^2) = \left\{ f/g \in \mathcal{O}_{(0,0)}(\mathbb{A}^2) \,:\, f(0,0) = 0 \right\}$, que es el conjunto de los $f/g \in \mathcal{O}_{(0,0)}(\mathbb{A}^2)$ tales que f no tiene término constante. Por tanto, $\mathfrak{m}_{(0,0)}(\mathbb{A}^2) = (x,y)$ y, por la propiedad (iii) de la proposición anterior, $\mu_{(0,0)}(x,y) = 1$.

La siguiente proposición nos proporciona dos técnicas que serán de utilidad a la hora de calcular la multiplicidad de la intersección de curvas proyectivas en un punto.

Proposición 1.35. *Sean C_f, C_g y C_h curvas planas afines y $P \in \mathbb{A}^2$. Se cumplen las siguientes propiedades:*

1. $\mu_P(f,g) = \mu_P(f, g+fh)$.
2. $\mu_P(f,gh) = \mu_P(f,g) + \mu_P(f,h)$.

Demostración. 1. Sigue de que $(f,g) = (f, g+fh)$.

2. Si C_f y C_g tienen alguna componente irreducible en común que contiene a P, entonces C_f y C_{gh} también la tienen, y por tanto $\mu_P(f,gh) = \infty = \mu_P(f,g) + \mu_P(f,h)$.

   Supongamos entonces que C_f y C_g no tienen componentes irreducibles en común que contengan a P. Dado que $(f,g) \subset (f,gh)$, podemos considerar el homomorfismo canónico $\pi : \mathcal{O}_P(\mathbb{A}^2)/(f,gh) \longrightarrow \mathcal{O}_P(\mathbb{A}^2)/(f,g)$. Sea además $\phi : \mathcal{O}_P(\mathbb{A}^2)/(f,h) \longrightarrow \mathcal{O}_P(\mathbb{A}^2)/(f,gh)$ la aplicación K-lineal dada por $\phi(\overline{z}) = \overline{gz}$ para $\overline{z} \in \mathcal{O}_P(\mathbb{A}^2)/(f,h)$. Comprobemos que

   $$0 \longrightarrow \mathcal{O}_P(\mathbb{A}^2)/(f,h) \xrightarrow{\phi} \mathcal{O}_P(\mathbb{A}^2)/(f,gh) \xrightarrow{\pi} \mathcal{O}_P(\mathbb{A}^2)/(f,g) \longrightarrow 0$$

   es una sucesión exacta corta.

   Es claro que π es sobreyectiva e im $\phi = \ker \pi$.

   Veamos que ϕ es inyectiva. Sea $\overline{z} \in \mathcal{O}_P(\mathbb{A}^2)/(f,h)$ tal que $\phi(\overline{z}) = \overline{gz} = 0$. Entonces $gz = uf + vgh$ para ciertos $u, v \in \mathcal{O}_P(\mathbb{A}^2)$. Dado que podemos multiplicar por un $s \in K[x,y]$ tal que $s(P) \neq 0$ para eliminar denominadores, podemos suponer que $u, v, z \in K[x,y]$. Entonces $g(z - vh) = uf$ y, como f y g no tienen factores en común, $f \mid z - vh$. Por tanto, $z - vh = wf$ para cierto $w \in K[x,y]$, lo que implica que $z = wf + vh$, de modo que $\overline{z} = 0 \in \mathcal{O}_P(\mathbb{A}^2)/(f,h)$.

   Por consiguiente, la sucesión es exacta y, tomando dimensiones, obtenemos que $\mu_P(f,gh) = \mu_P(f,g) + \mu_P(f,h)$. □

Ejemplo 1.36. Consideremos las curvas $C_f : y^2 - x^3 = 0$ y $C_g : x^2 - y^3 = 0$. Calculemos la multiplicidad de la intersección de C_f y C_g en el punto $P = (0,0)$.

En primer lugar, notemos que $P \in C_f \cap C_g$, de modo que $\mu_P(f,g) > 0$.

Usando las propiedades 1. y 2. de la proposición 1.35, se tiene que:

$$\begin{aligned}\mu_P(f,g) = \mu_P(y^2 - x^3, x^2 - y^3) &\overset{1.}{=} \mu_P(y^2 - x^3 + x(x^2 - y^3), x^2 - y^3) \\ &= \mu_P(y(y - xy^2), x^2 - y^3) \overset{2.}{=} \mu_P(y, x^2 - y^3) + \mu_P(y - xy^2, x^2 - y^3) \\ &\overset{2.}{=} \mu_P(y, x^2 - y^3) + \mu_P(y, x^2 - y^3) + \mu_P(1 - xy, x^2 - y^3) = 2 + 2 + 0 = 4,\end{aligned}$$

usando, en la penúltima desigualdad, que $\mu_P(y, x^2 - y^3) \overset{1.}{=} \mu_P(y, x^2) \overset{2.}{=} 2\,\mu_P(y,x) = 2$, por el ejemplo 1.34, y que $\mu_P(1 - xy, x^2 - y^3) = 0$ porque la curva dada por $1 - xy = 0$ no pasa por P.

Ejemplo 1.37. Sea C_f una curva que no contiene la recta $y = 0$ como componente irreducible, esto es, tal que $y \nmid f$. Calculemos $\mu_P(f,y)$ para $P = (0,0)$.

Por la propiedad 1. de la proposición 1.35, podemos reemplazar f por $f(x,0) \in K[x]$. Puesto que $y \nmid f$, se tiene que $f(x,0)$ no es el polinomio nulo, y podemos escribir $f(x,0) = x^m g$ con $g \in K[x]$, $g(0) \neq 0$. Es decir, m es el mayor número natural para el cual x^m divide a $f(x,0)$. Entonces

$$\mu_P(f,y) \overset{1.}{=} \mu_P(f(x,0),y) = \mu_P(x^m g, y) \overset{2.}{=} m\mu_P(x,y) + \mu_P(g,y) = m,$$

usando en la última igualdad que $\mu_{(0,0)}(x,y) = 1$, como hemos visto en el ejemplo 1.34, y que $g(0) \neq 0$, que implica que $\mu_P(g,y) = 0$ por la propiedad (ii) de la proposición 1.35.

A continuación definimos la *multiplicidad* de un punto en una curva, que intuitivamente representa cuántas veces debe ser contado como un punto en la misma. Emplearemos la siguiente notación:

Notación 1.38. Sea $f \in K[x,y]$ un polinomio de grado d. La *parte homogénea i-ésima*, que denotamos por f_i, $i = 0, \dots, d$, es la suma de todos los términos de f de grado i. Así, f se puede expresar como la suma de polinomios homogéneos $f = f_0 + f_1 + \dots + f_d$.

Definición 1.39. Sea C_f una curva. Definimos la *multiplicidad de C_f en el origen*, $m_{(0,0)}(C_f)$, como el menor n tal que la parte homogénea n-ésima de f es no nula, i.e., $f_n \neq 0$.

Sea $P \in \mathbb{A}^2$. Sea T la traslación tal que $T(0,0) = P$. Definimos la *multiplicidad de C_f en P*, $m_P(C_f)$, como $m_{(0,0)}(C_{f \circ T})$.

Obsérvese que $m_P(C_f) > 0 \iff P \in F$. Además, es inmediato comprobar que P es un punto regular de C_f si y solo si $m_P(C_f) = 1$. Si $m_P(C_f) > 1$, P es un punto singular.

Ejemplo 1.40. Consideremos las curvas del ejemplo 1.36, $C_f : y^2 - x^3 = 0$ y $C_g : x^2 - y^3 = 0$, donde habíamos visto que $\mu_{(0,0)}(C_f, C_g) = 4$, pese a corresponder a un único punto. Esto se debe a la singularidad de las curvas C_f y C_g en el origen, pues sigue de la definición anterior que su respectiva multiplicidad en él es $m_{(0,0)}(C_f) = 2$ y $m_{(0,0)}(C_g) = 2$.

Las construcciones que hemos visto hasta ahora en esta sección se extienden con facilidad al caso proyectivo:

Definición 1.41. (i) Sea $P \in \mathbb{P}^2$. Definimos el *anillo local* de P en $\mathbb{P}^2$ como el conjunto

$$\mathcal{O}_P(\mathbb{P}^2) := \left\{ \frac{F}{G} : F, G \in K[X,Y,Z] \text{ homogéneos del mismo grado}, G(P) \neq 0 \right\} \subset K(X,Y,Z).$$

Emplearemos también la notación $\mathcal{O}_P$.

(ii) Definimos el *homomorfismo de evaluación* en P como la aplicación

$$\mathcal{O}_P \longrightarrow K, \quad \frac{F}{G} \longmapsto \frac{F(P)}{G(P)},$$

y denotamos su núcleo por $\mathfrak{m}_P(\mathbb{P}^2) := \{ F/G \in \mathcal{O}_P : F(P) = 0 \}$, o también por $\mathfrak{m}_P$.

Observación 1.42. (a) Para ver que el homomorfismo de evaluación está bien definido basta razonar como en el punto 2. de la observación 1.25.

(b) Análogamente al caso afín, se tiene que $\mathcal{O}_P$ es un subanillo de $K(X,Y,Z)$ formado por elementos de grado 0 en $K(X,Y,Z)$ (donde el grado de un elemento se define como el grado de su numerador menos el grado de su denominador) y que es un anillo local en sentido algebraico, con $\mathfrak{m}_P$ como único ideal maximal.

(c) Se puede probar que para todo $P = (x_0 : y_0 : 1)$, la aplicación $\mathcal{O}_{(x_0,y_0)}(\mathbb{A}^2) \longrightarrow \mathcal{O}_{(x_0:y_0:1)}(\mathbb{P}^2)$, $\frac{f}{g} \longmapsto \frac{\widehat{f}}{\widehat{G}}$, con $\widehat{f}$ y $\widehat{g}$ las respectivas homogeneizaciones de f y g con respecto a Z, es un isomorfismo de anillos. Por tanto, las propiedades locales de una curva proyectiva en un punto P están caracterizadas por las propiedades locales de la curva afín correspondiente en (x_0, y_0).

La definición de multiplicidad de la intersección de dos curvas proyectivas es más delicada que en el caso afín, ya que el hecho de que los polinomios homogéneos no sean elementos de $\mathcal{O}_P$ nos obliga a adoptar primero una nueva definición para el ideal en $\mathcal{O}_P$ generado por polinomios homogéneos:

Definición 1.43. (i) Sean $F_1, \dots, F_k \in K[X,Y,Z]$ homogéneos. El *ideal homogéneo generado por* $F_1, \dots, F_k$ *en* $\mathcal{O}_P$ es el ideal

$$(F_1, \dots, F_k)^{\text{hom}} = \Big\{ \frac{f_1}{g_1}F_1 + \dots + \frac{f_k}{g_k}F_k \,:\, f_i = 0 \text{ o } f_i, g_i \in K[X,Y,Z] \text{ homogéneos} \\ \text{con } g_i(P) \neq 0 \text{ y } \deg(f_iF_i) = \deg(g_i) \text{ para todo } i \Big\} \subset \mathcal{O}_P.$$

(ii) Sean $\mathbb{P}^2$ el plano proyectivo sobre un cuerpo K, C_F y C_G dos curvas planas proyectivas y $P \in \mathbb{P}^2$. Definimos la *multiplicidad de la intersección* de C_F y C_G en el punto P como

$$\mu_P(C_F, C_G) := \dim_K \left(\mathcal{O}_P/(F,G)^{\text{hom}}\right) \in \mathbb{N} \cup \{\infty\}.$$

Emplearemos también la notación $\mu_P(F,G)$.

Observación 1.44. Por la parte (c) de la observación 1.42, si $P = (x_0 : y_0 : 1)$, del isomorfismo de anillos $\mathcal{O}_{(x_0:y_0:1)}(\mathbb{P}^2) \cong \mathcal{O}_{(x_0,y_0)}(\mathbb{A}^2)$ resulta que $\mu_{(x_0:y_0:1)}(F,G) = \mu_{(x_0,y_0)}(\widehat{F},\widehat{G})$. Por tanto, podemos emplear los resultados ya desarrollados para el caso afín para calcular la multiplicidad de la intersección proyectiva. Para los puntos del infinito, basta elegir otra inmersión de $\mathbb{A}^2$ en $\mathbb{P}^2$ y realizar el mismo proceso. Por consiguiente, la multiplicidad de la intersección es independiente de la elección de la coordenada que marca la recta del infinito.

Definición 1.45. Sea C_F una curva proyectiva y $P = (x_0 : y_0 : 1) \in \mathbb{P}^2$. Definimos la *multiplicidad de* C_F *en* P como $m_P(C_F) := m_{(x_0,y_0)}(C_{\widehat{F}})$, tomado como en la definición 1.39.

Definición 1.46. Un punto $P \in \mathbb{P}^2$ es un *punto de inflexión* de una curva proyectiva C_F si

1. P es un punto regular de C_F, y
2. si G es la recta tangente proyectiva de C_F en P, entonces $\mu_P(F,G) > 2$.

La recta tangente proyectiva en un punto de inflexión se denomina *tangente de inflexión*.

Ejemplo 1.47. Consideremos una curva de la forma $E : Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3$ con $a_1, \dots, a_6 \in K$, y sea $O = (0 : 1 : 0) \in E$.

O no es un punto singular de E, ya que $\frac{\partial F}{\partial Z}(O) = 1$.

Como además $\frac{\partial F}{\partial X}(O) = \frac{\partial F}{\partial Y}(O) = 0$, sigue de la definición 1.21 que la recta tangente a O es la recta del infinito $Z = 0$. Además, la multiplicidad de la intersección de E y $Z = 0$ en O es tres. Para comprobarlo, basta calcularla empleando la parte afín de E que cumple $y = 1$, gracias al isomorfismo de anillos $\mathcal{O}_{(x_0:1:z_0)}(\mathbb{P}^2) \cong \mathcal{O}_{(x_0,z_0)}(\mathbb{A}^2)$:

$$\begin{aligned}&\mu_{(0:1:0)}(Z, Y^2Z + a_1XYZ + a_3YZ^2 - X^3 - a_2X^2Z - a_4XZ^2 - a_6Z^3)\\ &= \mu_{(0,0)}(Z, Z + a_1XZ + a_3Z^2 - X^3 - a_2X^2Z - a_4XZ^2 - a_6Z^3) = 3,\end{aligned}$$

donde la última igualdad sigue del razonamiento expuesto en el ejemplo 1.37. Por tanto, O es un punto de inflexión de E, y $Z = 0$ es su tangente de inflexión.

Teorema 1.48 (Teorema de Bézout)**.** *Sean K un cuerpo y sean F y G dos curvas en $\mathbb{P}^2(K)$ de grado m y n, respectivamente. Supongamos que F y G no tienen componentes irreducibles en común. Entonces, F y G se intersecan a lo sumo en mn puntos, contados con sus respectivas multiplicidades de intersección:*

$$\sum_{P \in F \cap G} \mu_P(F, G) \leq mn.$$

Además, si K es algebraicamente cerrado, entonces se cumple la igualdad.

Demostración. Puede consultarse en [9, Appendix A.4] o [1, Section 5.3]. □

El teorema de Bézout será fundamental para probar, en el teorema 3.3, la asociatividad de la operación de grupo en las curvas elípticas. Este es el punto más delicado de la demostración del teorema, y en la literatura se pueden encontrar pruebas con otros enfoques: empleando fórmulas explícitas de la operación [8], utilizando el teorema de Riemann-Roch [8] o el teorema fundamental de Max Noether [1], pruebas geométricas [10], etc.

Como consecuencia del teorema de Bézout y de las propiedades de la multiplicidad de una curva en un punto, tenemos el siguiente resultado:

Corolario 1.49. *Toda curva proyectiva no singular sobre $\overline{K}$ es irreducible.*

Demostración. Puede consultarse en [6, Corollary 6.7]. □

Capítulo 2

Curvas elípticas y la ecuación de Weierstrass

Las curvas elípticas son, después de las cónicas, las curvas más ampliamente estudiadas [6], y han dado lugar a un rico marco teórico con conexiones a diversas áreas, desde geometría hasta la teoría de números y la criptografía. En este capítulo expondremos su definición formal y exploraremos sus propiedades fundamentales desde el punto de vista de la geometría algebraica. A continuación, nos centraremos en su clasificación en términos de isomorfimos y en su relación con el invariante j.

La exposición de este capítulo sigue fundamentalmente el clásico libro de Silverman [8] y, en lo referente al invariante j, los libros de Kunz [6] y Garrity [2]. De forma secundaria, también se consultaron los libros de Knapp [5], Milne [7] y Silverman-Tate [9].

2.1. Definiciones básicas sobre curvas elípticas

Esta sección está dedicada a nociones fundamentales sobre curvas elípticas y a la ecuación de Weierstrass, culminando con la obtención de su forma reducida y el estudio de algunas propiedades del determinante asociado a la misma.

Definición 2.1. Una *curva elíptica* sobre K es un par (E, O), donde E es una curva proyectiva plana no singular sobre K de grado 3 y $O \in E$.

En la literatura es también común una definición alternativa en la que se impone que E sea una curva proyectiva no singular de género 1 (para más detalles sobre el género de una curva, véase [1, Section 8.3]). No obstante, para una curva proyectiva plana no singular de grado d, el género se puede

calcular como

$$g = \frac{(d-1)(d-2)}{2}$$

[7, 1]. Por tanto, si $d = 1$ o $d = 2$, se tiene que $g = 0$. Si $d = 3$, entonces $g = 1$.

Ambas definiciones son, de hecho, equivalentes. En efecto, si E es una curva proyectiva no singular de género 1, se puede construir, empleando el teorema de Riemann-Roch, un isomorfismo entre E y una curva proyectiva plana en la llamada *forma de Weierstrass*:

$$Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3, \quad a_1, \dots, a_6 \in K, \tag{2.1}$$

donde podemos tomar $O = (0 : 1 : 0)$, obteniendo un caso particular de nuestra definición. Los detalles de esta construcción se pueden encontrar en [8, Proposition 3.1] o en [7, Chapter II]. De hecho, algunos textos como [5, 7] definen una curva elíptica como una curva proyectiva plana no singular dada por la *ecuación de Weierstrass* (2.1).

Observación 2.2. En ocasiones se exige también que el punto O sea un punto de inflexión. Como hemos visto en el ejemplo 1.47, el punto $(0 : 1 : 0)$ es un punto de inflexión de la curva proyectiva dada por (2.1), y se puede dar un cambio de variables que lleve el punto O a $(0 : 1 : 0)$ [7, Proposition 1.2], por lo que de nuevo nos encontramos ante una definición equivalente a la dada en 2.1.

En base a estas consideraciones y sin pérdida de generalidad, podemos suponer siempre que E está dada por la ecuación (2.1), que $O = (0 : 1 : 0)$, y denotar a la curva elíptica (E, O) simplemente como E.

Notación 2.3. Sea L una extensión del cuerpo K y sea E una curva elíptica sobre K. Emplearemos la notación $E(L)$ para referirnos al conjunto de puntos con coordenadas en L que cumple la ecuación de Weierstrass (2.1) que define a E. Es decir,

$$E(L) := \{(x : y : z) \in \mathbb{P}^2(L) : y^2z + a_1xyz + a_3yz^2 = x^3 + a_2x^2z + a_4xz^2 + a_6z^3\}.$$

Aun cuando $L = K$, en ocasiones escribiremos $E(K)$ en lugar de E cuando queramos enfatizar el cuerpo donde están definidos los puntos de la curva elípitica.

Consideremos ahora una curva elíptica C en la forma de Weierstrass (2.1). Imponiendo $Z = 0$, vemos que el único punto del infinito de la curva es O. Si suponemos $Z \neq 0$, podemos aplicar el cambio de coordenadas $x := X/Z$, $y := Y/Z$ para así obtener la ecuación de Weierstrass en forma afín:

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, \tag{2.2}$$

cuyos ceros en K^2 se corresponden con los puntos de la curva C en $\mathbb{P}^2$ distintos de O.

Si suponemos que $\operatorname{char} K \neq 2$, podemos sustituir y por $(y - a_1x - a_3)/2$, resultando

$$\frac{1}{4}y^2 - \frac{1}{4}a_1^2x^2 - \frac{1}{2}a_1a_3x - \frac{1}{4}a_3^2 = x^3 + a_2x^2 + a_4x + a_6.$$

Definiendo $b_2 := a_1^2 + 4a_2$, $b_4 := 2a_4 + a_1a_3$ y $b_6 := a_3^2 + 4a_6$, podemos reescribir la ecuación anterior como

$$y^2 = 4x^3 + b_2x^2 + 2b_4x + b_6.$$

Si además suponemos char $K \neq 3$, podemos ahora sustituir x por $(x - 3b_2)/36$ e y por $y/108$ (dado que $36 = 2^2 3^2$ y $108 = 2^2 3^3$). Agrupando términos, se obtiene

$$y^2 = x^3 + 27(24b_4 - b_2^2)x + 54(b_2^3 - 36b_2b_4 + 216b_6).$$

Finalmente, tomando $c_4 := b_2^2 - 24b_4$ y $c_6 := -b_2^3 + 36b_2b_4 - 216b_6$, llegamos a una expresión más sencilla de C:

$$y^2 = x^3 - 27c_4x - 54c_6. \tag{2.3}$$

Con el objetivo de simplificar la definición que sigue a continuación, introducimos la notación $b_8 := a_1^2a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a_3^2 - a_4^2$.

Definición 2.4. El *discriminante* de la curva dada por la ecuación de Weierstrass 2.1 se define como

$$\Delta := -b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6 \in K.$$

Es sencillo comprobar que $4b_8 = b_2b_6 - b_4^2$. Sabiendo esto, se demuestra que $1728\Delta = c_4^3 - c_6^2$.

Mediante el razonamiento expuesto previamente y utilizando la ecuación (2.3), podemos suponer que si char $K \neq 2, 3$, una curva elíptica puede ser descrita por la ecuación

$$Y^2Z = X^3 + AXZ^2 + BZ^3, \tag{2.4}$$

que, tras el cambio de coordenadas $x = X/Z$ e $y = Y/Z$, se convierte en

$$y^2 = x^3 + Ax + B, \tag{2.5}$$

donde $A, B \in K$. Nótese que la curva determinada por (2.5) tiene como clausura proyectiva a la curva determinada por (2.4). Por tanto, (2.5) también determina la curva elíptica, a falta del único punto del infinito $O = (0 : 1 : 0)$.

Con la notación anterior, se tiene que $a_1 = a_2 = a_3 = 0$, $a_4 = A$ y $a_6 = B$. Por ende, $b_2 = 0$, $b_4 = 2A$, $b_6 = 4B$ y $b_8 = -A^2$, con lo que $c_4 = -48A$ y $c_6 = -864B$. Así, dado que $\Delta = (c_4^3 - c_6^2)/1728$, se llega a

$$\Delta = -16(4A^3 + 27B^2).$$

Este resultado conduce a la siguiente proposición:

Proposición 2.5. *Sea E una curva dada en la forma de Weierstrass 2.1. Se tiene que E es no singular si y solo si $\Delta \neq 0$.*

Demostración. Sea E una curva dada por la ecuación 2.1. Como hemos visto en el ejemplo 1.47, el punto $O = (0 : 1 : 0)$ nunca es singular, por lo que nos basta estudiar los puntos afines de E.

Para evitar una demostración excesivamente densa, supondremos que char $K \neq 2, 3$ (los casos restantes se pueden consultar en [8, Appendix A, Proposition 1.2]). Entonces, podemos considerar la forma reducida de E dada por $y^2 = x^3 + Ax + B$.

($\Longrightarrow$) Tomemos un punto $(x_0 : y_0 : 1) \in E$, que identificaremos con el punto afín $P = (x_0, y_0)$. Consideremos el polinomio $f(x, y) := y^2 - x^3 - Ax - B$. E es singular en P si y solo si

$$\frac{\partial f}{\partial x}(x_0, y_0) = -3x_0^2 - A = 0,$$
$$\frac{\partial f}{\partial y}(x_0, y_0) = 2y_0 = 0.$$

La primera ecuación implica que $A = -3x_0^2$, y la segunda que $y_0 = 0$. Sustituyendo en f, obtenemos que $B = 2x_0^3$. En tal caso, $4A^3 + 27B^2 = -108x_0^6 + 108x_0^6 = 0$, lo que implica que $\Delta = 0$.

($\Longleftarrow$) Supongamos ahora que $\Delta = 0$, esto es, que $4A^3 + 27B^2 = 0$. Entonces, el polinomio cúbico $g(x) := x^3 + Ax + B$ tiene una raíz doble, digamos x_0, pues $4A^3 + 27B^2$ es su determinante (para más detalles, consúltese el capítulo 4 de [11]). Esto ocurre si y solo si x_0 es también raíz de la derivada de g, $g'(x) = 3x^2 + A$. En tal caso, para el polinomio $f(x, y) = y^2 - x^3 - Ax - B$ tenemos

$$f(x_0, 0) = -x_0^3 - Ax_0 - B = 0,$$
$$\frac{\partial f}{\partial x}(x_0, 0) = -3x_0^2 - A = 0,$$
$$\frac{\partial f}{\partial y}(x_0, 0) = 0.$$

Por tanto, E es singular en $(x_0, 0)$. □

Observación 2.6. Consideremos el polinomio correspondiente a la ecuación de Weierstrass en forma afín, $f(x, y) = y^2 + a_1xy + a_3y - x^3 - a_2x^2 - a_4x - a_6$, con $a_1, \ldots, a_6 \in K$, y la curva $C : f = 0$. Si $P = (x_0, y_0) \in C$ es un punto singular de C, entonces existen $\alpha, \beta \in \overline{K}$ tales que la serie de Taylor de f centrada en (x_0, y_0) viene dada por

$$f(x, y) = ((y - y_0) - \alpha(x - x_0))((y - y_0) - \beta(x - x_0)) - (x - x_0)^3,$$

usando que $f(x_0, y_0) = \frac{\partial f}{\partial x}(x_0, y_0) = \frac{\partial f}{\partial y}(x_0, y_0) = 0$.

Definición 2.7. Con la notación de la observación precedente, se dice que el punto singular P es un *nodo* si $\alpha \neq \beta$. En tal caso, las rectas tangentes a C en P son los factores lineares de $((y - y_0) - \alpha(x - x_0))((y - y_0) - \beta(x - x_0))$, es decir, $y - y_0 = \alpha(x - x_0)$ e $y - y_0 = \beta(x - x_0)$.

Si $\alpha = \beta$, se dice que P es una cúspide. Habrá entonces una recta tangente ("doble"), dada por $y - y_0 = \alpha(x - x_0)$.

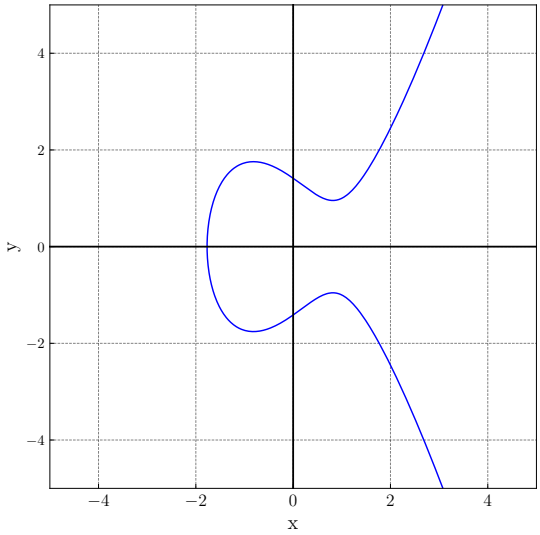


Figura 2.1: Curva elíptica definida sobre $\mathbb{R}$ y dada por $y^2 = x^3 - 2x + 2$, un caso particular de la forma (2.5).

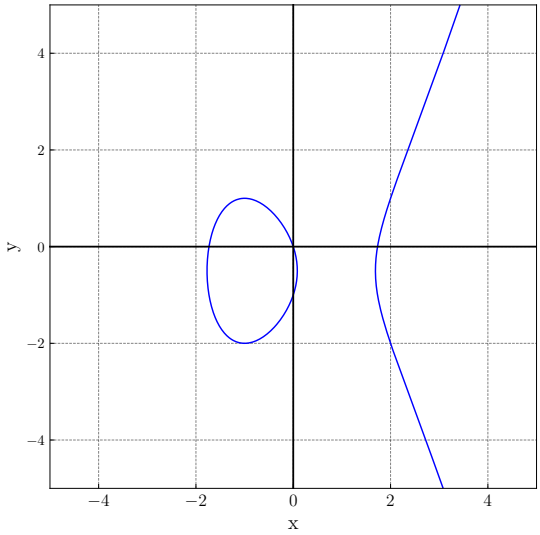


Figura 2.2: Curva elíptica definida sobre $\mathbb{R}$ y dada por $y^2 + y = x^3 - 3x$. sin simetría respecto al eje x.

Ejemplo 2.8. Para los siguientes ejemplos, tomaremos $K = \mathbb{R}$.

1. La curva elíptica $E_1 : y^2 = x^3 - 2x + 2$ tiene discriminante $\Delta = -1216 \neq 0$. Obsérvese en la Figura 2.1 la simetría con respecto al eje x, que es una propiedad de cualquier curva elíptica expresada en la forma (2.5).
2. La curva elíptica $E_2 : y^2 + y = x^3 - 3x$, con discriminante $\Delta = 1701 \neq 0$, no es simétrica respecto al eje x, como se ve en la Figura 2.2. No obstante, podría aplicarse un cambio de variables para imponer dicha simetría, dado que $\text{char}\,\mathbb{R} = 0 \neq 2, 3$.
3. La curva $C_f : f = 0$, con $f(x, y) = y^2 - x^3 + 3x - 2 \in \mathbb{R}[x, y]$, representada en la Figura 2.3, es singular. Dado que $\frac{\partial f}{\partial x}(x, y) = -3x^2 + 3$ y $\frac{\partial f}{\partial y}(x, y) = 2y$, el único punto singular de C_f es $P = (1, 0) \in C_f$. Como además $f(x, y) = y^2 - 3(x-1)^2 - (x-1)^3 = (y - \sqrt{3}(x-1))(y + \sqrt{3}(x-1)) - (x-1)^3$, se tiene que las rectas tangentes a P son $y = \sqrt{3}(x-1)$ e $y = -\sqrt{3}(x-1)$, y que P es un nodo.
4. La curva $C_g : g = 0$, con $g(x, y) = y^2 - x^3 \in \mathbb{R}[x, y]$, representada en la Figura 2.4, es singular. Dado que $\frac{\partial f}{\partial x}(x, y) = 3x^2$ y $\frac{\partial f}{\partial y}(x, y) = 2y$, el único punto singular de C_g es $Q = (0, 0) \in C_g$. En este caso, la única recta tangente a Q es $y = 0$ y Q es una cúspide.

Proposición 2.9. *Sea C una curva dada por la ecuación de Weierstrass* (2.1).

(a) C tiene un nodo si y solo si $\Delta = 0$ y $c_4 \neq 0$.

(b) C tiene una cúspide si y solo si $\Delta = c_4 = 0$.

Si (a) o (b) se cumplen, existe un único punto singular.

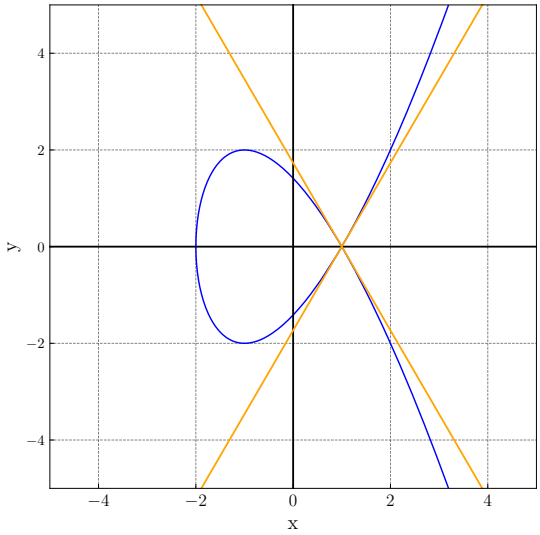


Figura 2.3: En azul, la curva definida sobre $\mathbb{R}$ dada por la ecuación $y^2 = x^3 - 3x + 2$. En naranja, las rectas tangentes al nodo $(1, 0)$.

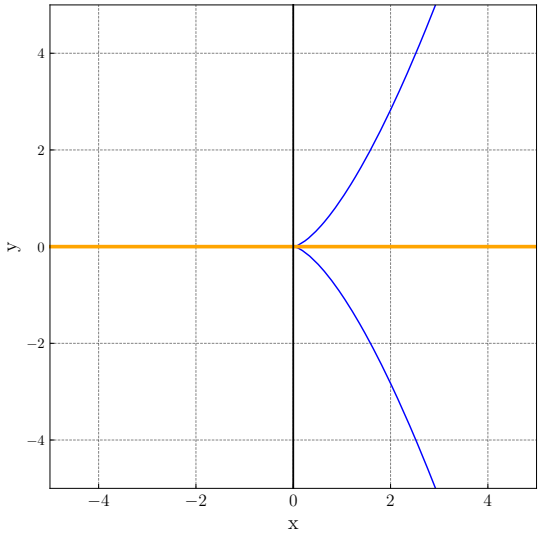


Figura 2.4: En azul, la curva definida sobre $\mathbb{R}$ dada por la ecuación $y^2 = x^3$. En naranja, la recta tangente (doble) a la cúspide $(0, 0)$.

Demostración. Sea C la curva cuya forma afín está dada por $C : f(x, y) = y^2 + a_1xy + a_3y - x^3 - a_2x^2 - a_4x - a_6 = 0$, con $a_1, \dots, a_6 \in K$.

Supongamos que $\Delta = 0$. Equivalentemente, por la proposición 2.5, C es singular.

Como se vio en el ejemplo 1.47, el punto en el infinito es regular, por lo que debe existir un punto afín $P \in C$ singular. Sea $P = (x_0, y_0) \in \mathbb{A}^2$.

Se puede comprobar que el cambio de variables $x = x' + x_0$, $y = y' + y_0$ no cambia Δ ni c_4, por lo que podemos asumir, sin pérdida de generalidad, que $P = (0, 0)$. Entonces

$$a_6 = -f(0,0) = 0, \quad a_4 = -\frac{\partial f}{\partial x}(0,0) = 0, \quad a_3 = \frac{\partial f}{\partial y}(0,0) = 0$$

y C toma la expresión $C : y^2 + a_1xy - a_2x^2 - x^3 = 0$. Por tanto, $b_2 = a_1^2 + 4a_2$ y $b_4 = 0$, con lo cual $c_4 = b_2^2 = (a_1^2 + 4a_2)^2$.

Por la definición 2.7, C tendrá un nodo en P si la forma cuadrática $y^2 + a_1xy - a_2x^2$ tiene factores distintos, y una cúspide si son iguales. Dado que

$$y^2 + a_1xy - a_2x^2 = (y + \frac{a_1}{2}x)^2 - \frac{a_1^2}{4}x^2 - a_2x^2 = (y + \frac{a_1}{2}x)^2 - \frac{a_1^2 + 4a_2}{4}x^2,$$

C tendrá un nodo si el discriminante de esta forma cuadrática, $a_1^2 + 4a_2$, es distinto de 0, y una cúspide si es nulo.

Puesto que los razonamientos usados son igualmente válidos en la dirección opuesta, los apartados (a) y (b) quedan demostrados.

Para probar la última afirmación de la proposición, asumiremos char $K \neq 2$ (el caso char $K = 2$ se puede consultar en [8, Appendix A]). En ese caso, sabemos que la ecuación de C se puede escribir como $y^2 = 4x^3 + b_2x^2 + 2b_4x + b_6$. Derivando, tenemos que un punto $(x_0, y_0) \in C$ es singular si y solo si $2y_0 = 12x^2 + 2b_2x + 2b_4 = 0$. Es decir, si y solo si $y_0 = 0$ y x_0 es una raíz doble del polinomio cúbico $4x^3 + b_2x^2 + 2b_4x + b_6$. Puesto que un polinomio cúbico no puede tener dos raíces dobles, a lo sumo existe un punto singular. □

Observación 2.10. Este resultado nos permitiría haber deducido que las curvas C_f y C_g del ejemplo 2.8 tienen un nodo y una cúspide en $(1, 0)$ y $(0, 0)$, respectivamente, aunque no nos proporciona la ecuación de las rectas tangentes directamente.

Observación 2.11. Puede demostrarse que los resultados de la proposición 2.9 tienen la siguiente generalización: una curva cúbica irreducible tiene a lo sumo un punto singular, que es o bien un nodo o bien una cúspide [1, Chapter 5]. Estas pueden ser consideradas formas degeneradas de las curvas elípticas.

2.2. Aplicaciones entre curvas elípticas

Las aplicaciones entre curvas elípticas, que reciben el nombre de *isogenias*, juegan un importante papel en el estudio de las curvas elípticas y han dado lugar a ricas y profundas teorías. En el ámbito teórico, destaca por ejemplo su conexión con los módulos de Tate. En el ámbito práctico, es relevante su uso para la construcción de algoritmos criptográficos. Por ejemplo, el protocolo conocido como *supersingular isogeny Diffie Hellman* (SIDH) fue uno de los candidatos más prometedores para el desarrollo de criptografía post-cuántica hasta que una vulnerabilidad insalvable fue encontrada en 2022 [12].

Por motivos de espacio, nos limitaremos a dar una breve introducción a estos objetos y a enunciar, sin demostración, algunas de sus propiedades fundamentales. Para un desarrollo exhaustivo, sugerimos consultar [8].

Definición 2.12. Sean E_1 y E_2 curvas elípticas sobre K. Una *isogenia* de E_1 a E_2 es un morfismo de curvas $\phi : E_1(\overline{K}) \to E_2(\overline{K})$ tal que $\phi(O) = O$.

Todo morfismo de curvas es constante o sobreyectivo [18, Proposition II.6.8], de modo que una isogenia puede ser de dos tipos: o bien $\phi(E_1(\overline{K})) = \{O\}$, o bien $\phi(E_1(\overline{K})) = E_2(\overline{K})$. Si ϕ es constante, la llamaremos *isogenia nula* y la denotaremos por $[0]$.

Definición 2.13. Dos curvas elípticas sobre K, E_1 y E_2, son *isomorfas* sobre $\overline{K}$ si existen isogenias $\phi : E_1(\overline{K}) \to E_2(\overline{K})$ y $\psi : E_2(\overline{K}) \to E_1(\overline{K})$ tales que las composiciones $\phi \circ \psi$ y $\psi \circ \phi$ son la identidad. En tal caso, ϕ y ψ son *isomorfismos* (de curvas elípticas).

Si ϕ y ψ se pueden definir sobre $E_1(K)$ y $E_2(K)$ de forma que $\phi\circ\psi$ y $\psi\circ\phi$ sean la identidad, diremos que E_1 y E_2 son *isomorfas sobre* K.

Nos interesarán isomorfismos específicos: aquellos dados por *cambios de variables admisibles*, que son cambios de coordenadas de la forma

$$x = u^2x' + r, \quad y = u^3y' + u^2sx' + t,$$

con $u, r, s, t \in \overline{K}$ y $u \neq 0$. Estos se caracterizan por fijar el punto $O = (0 : 1 : 0)$, llevar la recta del infinito a sí misma, y mantener la forma de Weierstrass (2.2). A salvo de una constante, los cambios de variables admisibles son las transformaciones lineales más generales con estas propiedades [5, Chapter 2].

Para curvas elípticas de la forma $E : y^2 = x^3 + Ax + B$ y $E' : y^2 = x^3 + A'x + B'$, con $A, A', B, B' \in K$, tenemos un resultado más fuerte:

Lema 2.14. *Sea char* $K \neq 2, 3$. *Sean* $E : y^2 = x^3 + Ax + B$ *y* $E' : y^2 = x^3 + A'x + B'$ *curvas elípticas sobre* K *y* $\phi : E'(\overline{K}) \to E(\overline{K})$ *un isomorfismo tal que* $\phi(O) = O$. *Entonces existe* $u \in \overline{K}^\times$ *tal que* $A' = u^4A$, $B' = u^6B$ *y* ϕ *está dado por* $\phi(x : y : z) = (u^2x : u^3y : z)$.

Recíprocamente, si $A' = u^4A$, $B' = u^6B$ *y* $\phi(x : y : z) = (u^2x : u^3y : z)$ *para un cierto* $u \in \overline{K}^\times$, *entonces* ϕ *es un isomorfismo de* E' *en* E *tal que* $\phi(O) = O$.

Demostración. Véase [7, Theorem II.2.1]. □

2.3. El invariante j

A lo largo de esta sección, supondremos char $K \neq 2$.

Proposición 2.15. *Toda curva elíptica* E *sobre* $\overline{K}$ *es isomorfa a una curva elíptica dada por*

$$E_\lambda : y^2 = x(x-1)(x-\lambda), \tag{2.6}$$

donde $\lambda \in \overline{K}$ *y* $\lambda \neq 0, 1$.

Demostración. Como hemos visto en la sección 2.1, dado que char $\overline{K} \neq 2$, podemos suponer que E está dada por una ecuación de Weierstrass en la forma $y^2 = 4x^3 + b_2x^2 + 2b_4x + b_6$, con $b_2, b_4, b_6 \in \overline{K}$.

Reemplazando (x, y) por $(x, 2y)$, se obtiene una ecuación donde el polinomio cúbico del lado derecho es mónico. Factorizándolo, podemos reescribir la ecuación de E como

$$y^2 = (x-\alpha)(x-\beta)(x-\gamma), \tag{2.7}$$

con $\alpha, \beta, \gamma \in \overline{K}$.

Por definición, el discriminante del polinomio en una variable $(x-\alpha)(x-\beta)(x-\gamma)$ es $d := (\alpha-\beta)^2(\alpha-\gamma)^2(\beta-\gamma)^2$ [5, p. 59]. Sabiendo que d es invariante bajo transformaciones lineales de coordenadas y que la relación entre este y el discriminante de E es $\Delta = 16d$ [5, Proposition 3.6], tenemos en este caso que $\Delta = 16(\alpha-\beta)^2(\alpha-\gamma)^2(\beta-\gamma)^2$. Por la proposición 2.5, E es no singular si y solo si $\Delta \neq 0$, lo que implica que α, β y γ son distintos dos a dos.

Realizando el cambio de variables $x = (\beta-\alpha)u + \alpha$, $y = (\beta-\alpha)^{3/2}v$ y sustituyendo en (2.7), se obtiene

$$v^2 = u(u-1)(u-\lambda), \text{ con } \lambda = \frac{\gamma-\alpha}{\beta-\alpha} \in \overline{K},$$

verificándose que $\lambda \neq 0, 1$. □

Definición 2.16. Se dice que una curva elíptica dada por la ecuación (2.6) está en *forma de Legendre*.

Observación 2.17. Si $\lambda = 0$ o $\lambda = 1$, la curva dada por la ecuación $y^2 = x(x-1)(x-\lambda)$ es singular en el punto $(\lambda, 0)$ (lo cual se ve fácilmente tomando derivadas parciales), por lo que no define una curva elíptica.

En la ecuación (2.7) hemos factorizado el polinomio cúbico del lado derecho, y definido λ en función de sus raíces α, β y γ. Realizar una permutación de estas raíces no cambia ni el polinomio ni la curva elíptica resultantes, pero λ sí se ve afectado. Por ejemplo, al permutar γ y β obtendríamos $1/\lambda$ en lugar de λ. Esto implica que las curvas elípticas dadas por $x(x-1)(x-\lambda)$ y $x(x-1)(x-1/\lambda)$ deberían ser equivalentes. Calculando el resultado para cada una de las seis permutaciones de un conjunto de tres elementos $\{1, 2, 3\}$, $S_3 = \{(1), (12), (13), (23), (123), (132)\}$, tenemos

$$(1)\quad \frac{\gamma-\alpha}{\beta-\alpha} = \lambda \qquad (12)\quad \frac{\gamma-\beta}{\alpha-\beta} = 1-\lambda$$

$$(13)\quad \frac{\alpha-\gamma}{\beta-\gamma} = \frac{\lambda}{\lambda-1} \qquad (23)\quad \frac{\beta-\alpha}{\gamma-\alpha} = \frac{1}{\lambda}$$

$$(123)\quad \frac{\alpha-\beta}{\gamma-\beta} = \frac{1}{1-\lambda} \qquad (132)\quad \frac{\beta-\gamma}{\alpha-\gamma} = \frac{\lambda-1}{\lambda}.$$

Existe por tanto una correspondencia seis-a-uno entre los posibles valores de λ en la ecuación (2.6) y la curva elíptica resultante (a menos que algunos de estos valores coincidan, como explicaremos en la observación 2.22). En otras palabras, una curva elíptica no está unívocamente caracterizada por λ. En su lugar, querríamos poder dar un valor que pudiese ser definido a partir de la ecuación (2.6) y que fuese invariante frente a la elección tomada en el conjunto

$$M_\lambda := \left\{\lambda, 1-\lambda, \frac{\lambda}{\lambda-1}, \frac{1}{\lambda}, \frac{1}{1-\lambda}, \frac{\lambda-1}{\lambda}\right\}.$$

Este valor recibe el nombre de *invariante j*.

Definición 2.18. Consideremos una curva elíptica en forma de Legendre $E_\lambda : y^2 = x(x-1)(x-\lambda)$, con $\lambda \in \overline{K}$, $\lambda \neq 0, 1$. El *invariante j* de E_λ es el valor

$$j(E_\lambda) = 2^8 \frac{(\lambda^2 - \lambda + 1)^3}{\lambda^2(\lambda - 1)^2}. \tag{2.8}$$

Proposición 2.19. *Sea $\lambda \in \overline{K}$, $\lambda \neq 0, 1$. Las curvas elípticas E_λ, $E_{1-\lambda}$, $E_{\frac{\lambda}{\lambda-1}}$, $E_{\frac{1}{\lambda}}$, $E_{\frac{1}{1-\lambda}}$ y $E_{\frac{\lambda-1}{\lambda}}$ tienen el mismo invariante j.*

Demostración. Veamos que $j(E_{1-\lambda}) = j(E_\lambda)$.

$$j(E_{1-\lambda}) = 2^8 \frac{((1-\lambda)^2 - (1-\lambda) + 1)^3}{(1-\lambda)^2((1-\lambda)-1)^2} = 2^8 \frac{(\lambda^2 - \lambda + 1)^3}{(1-\lambda)^2\lambda^2} = j(E_\lambda).$$

El resto de comprobaciones se realizan fácilmente con cálculos análogos. □

La siguiente proposición refina el resultado anterior, pues nos dice que ninguna otra curva elíptica distinta de E_λ, $E_{1-\lambda}$, $E_{\frac{\lambda}{\lambda-1}}$, $E_{\frac{1}{\lambda}}$, $E_{\frac{1}{1-\lambda}}$ y $E_{\frac{\lambda-1}{\lambda}}$ tiene el mismo invariante j que estas.

Proposición 2.20. *Supongamos que*

$$a = 2^8 \frac{(\mu^2 - \mu + 1)^3}{\mu^2(\mu - 1)^2}$$

para una cierta constante a y $\mu \in \overline{K}$, $\mu \neq 0, 1$. Entonces, un valor λ es solución de la ecuación $2^8(\mu^2 - \mu + 1)^3 - a\mu^2(\mu - 1)^2 = 0$ si y solo si $j(E_\lambda) = a$, y existen otras cinco soluciones de la ecuación: $1 - \lambda$, $\lambda/(\lambda - 1)$, $1/\lambda$, $1/(1 - \lambda)$ y $(\lambda - 1)/\lambda$.

Demostración. Si λ es una solución, entonces $2^8(\lambda^2 - \lambda + 1)^3 - a\lambda^2(\lambda - 1)^2 = 0$. Despejando a obtenemos que

$$a = 2^8 \frac{(\lambda^2 - \lambda + 1)^3}{\lambda^2(\lambda - 1)^2} = j(E_\lambda),$$

de donde se ve también la implicación en el otro sentido.

El grado de $2^8(\mu^2 - \mu + 1)^3 - a\mu^2(\mu - 1)^2 = 0$ respecto a μ es 6, de modo que, por el teorema fundamental del álgebra, existen exactamente 6 soluciones. Por la proposición 2.19, los valores $1-\lambda$, $\lambda/(\lambda - 1)$, $1/\lambda$, $1/(1 - \lambda)$ y $(\lambda - 1)/\lambda$ definen sendas formas de Legendre con el mismo invariante j, por lo que deben ser las soluciones restantes. □

Corolario 2.21. *Dos curvas elípticas son isomorfas sobre $\overline{K}$ si y solo si su invariante j coincide.*

Demostración. Sigue de la proposición 2.20. □

Observación 2.22. Es posible que dos o más valores del conjunto M_λ coincidan. Igualándolos dos a dos, se ve que esto ocurre cuando $\lambda \in \{-1, 2, 1/2\}$ o $\lambda^2 - \lambda + 1 = 0$. En el primer caso, en lugar de seis-a-uno, la asociación $\lambda \mapsto j(E_\lambda)$ será tres-a-uno. En el segundo, será dos-a-uno. Estas dos posibilidades se corresponden con los invariantes j 1728 y 0, respectivamente. Nótese que cuando char $K = 3$, entonces $1728 = 0$ y hay un único caso posible. Se tiene entonces que M_λ solo contiene un elemento, $\lambda = -1$.

El invariante j tiene también un significado geométrico, pues es invariante a través de las transformaciones proyectivas que llevan una curva elíptica en otra. Por consiguiente, el invariante j clasifica las curvas elípticas. La demostración de esta propiedad requiere la introducción de herramientas adicionales del ámbito de la geometría proyectiva, como las transformaciones proyectivas y la noción de *razón doble*, por lo que remitimos al lector al anexo I para un desarrollo completo.

Una pregunta natural es cuál es la relación entre la expresión (2.8) del invariante j, dada a partir de la forma de Legendre de una curva elíptica, y la forma de Weierstrass de la misma. Dado que es más habitual trabajar con la forma de Weierstrass que con la forma de Legendre, estamos interesados en encontrar una expresión del invariante j que solamente dependa de los coeficientes de la primera.

Para facilitar la exposición, daremos primero la expresión asociada a la fórmula de Weierstrass del invariante j y demostraremos a continuación que esta es equivalente a la de la definición 2.18 para la fórmula de Legendre.

Proposición 2.23. *Sea $E_\lambda : y^2 = x(x-1)(x-\lambda)$ una curva elíptica sobre $\overline{K}$ con forma de Weierstrass $y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$, para ciertos $a_1, \dots, a_6 \in \overline{K}$. El invariante j de E_λ está dado por*

$$j(E_\lambda) = \frac{c_4^3}{\Delta} = \frac{c_4^3}{-b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6}. \tag{2.9}$$

Demostración. Comencemos transformando la forma de Legendre en la forma de Weierstrass. Observamos que

$$y^2 = x(x-1)(x-\lambda) = x^3 - x^2 - \lambda x^2 + x\lambda = x^3 - (1+\lambda)x^2 + \lambda x.$$

Tenemos entonces que $a_1 = a_3 = a_6 = 0$, $a_2 = -(1+\lambda)$ y $a_4 = \lambda$. Por consiguiente, $b_2 = -4(1+\lambda)$, $b_4 = 2\lambda$, $b_6 = 0$ y $b_8 = -\lambda^2$, de modo que $c_4 = 16(1+\lambda)^2 - 48\lambda = 16(\lambda^2 - \lambda + 1)$ y $c_6 =$

$64(1+\lambda)^3 - 288(1+\lambda)\lambda = 32(1+\lambda)(2\lambda^2 - 5\lambda + 2)$. Así,

$$\begin{aligned}\frac{c_4^3}{\Delta} &= 1728\,\frac{16^3(\lambda^2-\lambda+1)^3}{16^3(\lambda^2-\lambda+1)^3 - 32^2(1+\lambda)^2(2\lambda^2-5\lambda+2)^2} \\ &= 1728\,\frac{16^3(\lambda^2-\lambda+1)^3}{27648\lambda^4 - 55296\lambda^3 + 27648\lambda^2} \\ &= 1728\,\frac{16^3(\lambda^2-\lambda+1)^3}{27648\lambda^2(\lambda^2-2\lambda+1)} \\ &= 2^8\,\frac{(\lambda^2-\lambda+1)}{\lambda^2(\lambda-1)^2} = j(E_\lambda).\end{aligned}$$

□

Observación 2.24.

1. La fórmula del invariante j (2.9) es válida aunque char $\overline{K} = 2$, y también en este caso se puede demostrar que el invariante j clasifica las curvas elípticas [8, Proposition III.1.4].
2. Si char $\overline{K} \neq 2, 3$ y consideramos una curva elíptica $E : y^2 = x^3 + Ax + B$, se tiene que [8, Appendix A.1.2]
$$j = -1728\frac{(4A)^3}{\Delta} = -1728\frac{(4A)^3}{4A^3 + 27B^2}. \tag{2.10}$$
3. Aunque tanto las fórmulas (2.9) como (2.10) son válidas para curvas elípticas definidas sobre un cuerpo $K \subsetneq \overline{K}$, no es cierto en general que curvas elípticas sobre cuerpos no algebraicamente cerrados con el mismo invariante j sean isomorfas sobre estos. Por ejemplo, las curvas $E : y^2 = x^3 + ax + b$ y $E_d : y^2 = x^3 + ad^2x + bd^3$, con $d \in K^\times$ un entero libre de cuadrados, tienen el mismo invariante $j = -1728\frac{(4a)^3}{4a^3+27b^2}$, pero basta usar el lema 2.14 para ver que no son isomorfas. No obstante, sí lo serían en $K(\sqrt{d})$ y, por tanto, en $\overline{K}$. Este es un ejemplo de lo que se conoce como *forma torcida* de una curva elíptica.

Proposición 2.25. *Sea $j_0 \in \overline{K}$. Existe una curva elíptica E tal que $j(E) = j_0$.*

Demostración. Si $j_0 = 0$, podemos tomar $E_1 : y^2 + y = x^3$, cuyo discriminante es $\Delta = -27$.

Si $j_0 = 1728$, tomamos $E_2 : y^2 = x^3 + x$, cuyo discriminante es $\Delta = -64$, pues $b_2 = b_4 = b_6 = 0$ y $b_8 = -1$, con la notación de la sección 2.1.

Nótese que si char K es 2 o 3, se tiene que $1728 = 0$ y basta elegir, de entre E_1 y E_2, la curva que sea no singular.

Si $j_0 \neq 0, 1728$, podemos tomar

$$E : y^2 + xy = x^3 - \frac{36}{j_0 - 1728}x - \frac{1}{j_0 - 1728}$$

que cumple $c_4 = \frac{j_0}{j_0-1728}$ y $c_6 = \frac{-j_0}{j_0-1728}$, por lo que $\Delta = \frac{j_0^2}{(j_0-1728)^3}$. □

Capítulo 3

La estructura de grupo en curvas elípticas

Una de las características más destacadas de las curvas elípticas es el hecho de que se las puede dotar de una estructura de grupo abeliano. Este tiene a su vez una interpretación geométrica, pues la operación de grupo, en su formulación más básica, consiste en unir dos puntos de la curva por una recta y reflejar el tercer punto de intersección de la recta y la curva elíptica sobre el eje x (de ahí que se pueda encontrar referido en algunos textos como el *chord-and-tangent method*[1]).

El teorema de Mordell establece que el grupo de puntos racionales de una curva elíptica es finitamente generado, y el teorema de Lutz-Nagell caracteriza explícitamente los puntos de torsión racionales. Sin embargo, aún no se dispone de un criterio general para determinar el rango de la parte libre de torsión del grupo, y se conjetura que puede ser arbitrariamente grande.

En este capítulo definiremos la ley de grupo y probaremos que esta da lugar a un grupo abeliano. A continuación, expondremos algunos de los principales resultados al respecto. Como referencias, recomendamos los libros de Silverman [8], Silverman y Tate [9], Kunz [6] y Knapp [5].

3.1. Construcción

Sea (E, O) una curva elíptica definida sobre un cuerpo algebraicamente cerrado $\overline{K}$, con $O \in E$ arbitrario, y sea $L \subset \mathbb{P}^2(\overline{K})$ una recta proyectiva. Por el teorema de Bézout (teorema 1.48), L y E se intersecan en exactamente tres puntos, contando multiplicidades (i.e., dichos puntos pueden no ser distintos). Teniendo en cuenta este hecho, definimos la *ley de composición* o *ley de grupo* como sigue:

[1] Se puede traducir *chord-and-tangent method* como "método de la cuerda y la tangente"

Definición 3.1. Sean $P, Q \in E$ y sea L la recta que pasa por P y Q (la recta tangente a E en P, si $P = Q$). Denotamos por $P * Q$ el tercer punto de intersección de L y E.

Sea ahora L' la recta que pasa por $P * Q$ y por un punto fijado $O \in E$. Definimos como $P + Q$ el tercer punto de intersección de L' y E, esto es,

$$P + Q := (P * Q) * O.$$

Veremos que esta definición, de la cual se da un ejemplo en la Figura 3.1, nos da una estructura de grupo abeliano en E. Para ello, utilizaremos el siguiente lema:

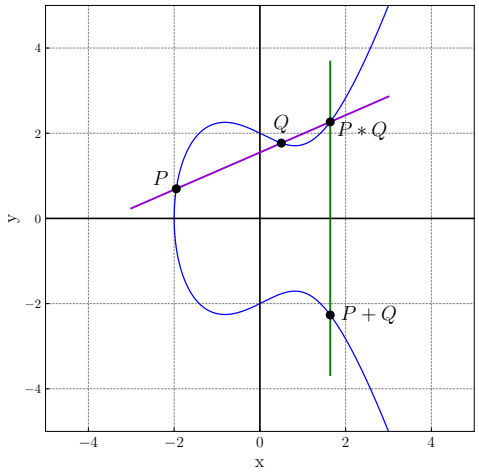


Figura 3.1: Suma de dos puntos sobre la curva elíptica dada por la ecuación $y^2 = x^3 - 2x + 4$, tomando $O = (0 : 1 : 0)$.

Lema 3.2. *Sean C, C_1 y C_2 curvas cúbicas proyectivas. Si C contiene ocho puntos de intersección de C_1 y C_2, entonces C contiene el noveno punto de intersección de C_1 y C_2.*

Demostración. En primer lugar, nótese que, como consecuencia del teorema de Bézout, dos curvas cúbicas tienen exactamente nueve puntos de intersección, contando multiplicidades.

Una curva cúbica proyectiva está dada por una ecuación de la forma $F(X, Y, Z) = 0$, donde F es un polinomio homogéneo de grado tres, y que por tanto se puede escribir como $\sum a_{ij} X^i Y^j Z^{3-i-j} = 0$, con $0 \leq i, j \leq 3$ e $i + j \leq 3$. Entonces, una curva cúbica está determinada por diez coeficientes a_{ij}.

Imponer que la curva pase por un punto dado equivale a imponer una condición lineal en los coeficientes a_{ij}. Por consiguiente, imponer que pase por ocho puntos $P_1, \ldots, P_8 \in \mathbb{P}^2$ resulta en un espacio de $10 - 8 = 2$ dimensiones.

Supongamos que C_1 y C_2 están dadas por las ecuaciones $C_1 : F = 0$ y $C_2 : G = 0$, con $F, G \in K[X, Y, Z]$ homogéneos de grado tres. Entonces, dados $\lambda_1, \lambda_2 \in K^\times$, se tiene que $\lambda_1 F + \lambda_2 G$ define

una curva cúbica que pasa por $P_1, \dots, P_8$. Como $\lambda_1 F + \lambda_2 G$ es un espacio de dimensión dos, debe ser exactamente el espacio de las curvas cúbicas que pasan por $P_1, \dots, P_8$ y, por tanto, C debe estar dada por una ecuación de la forma $\lambda_1 F + \lambda_2 G = 0$. Esto implica que, como $F(P_9) = G(P_9) = 0$, C contiene a P_9. □

Teorema 3.3. *El par $(E, +)$, siendo $+$ la ley de composición*

$$\begin{aligned} + : E \times E &\longrightarrow E \\ (P, Q) &\longmapsto P + Q \end{aligned}$$

es un grupo abeliano con O como elemento neutro. Denotaremos por $-P$ el elemento opuesto a un $P \in E$ dado.

Demostración. O es, en efecto, el elemento neutro, pues dado $P \in E$, tenemos que $P + O = (P * O) * O = P$, donde la última igualdad sigue de que el tercer punto de intersección con E de la recta que pasa por $P * O$ y O es P, por definición de $P * O$.

Veamos que todo $P \in E$ tiene elemento opuesto. Sea $-P = P * (O * O)$. Entonces, el tercer punto de intersección con E de la recta que une P y $-P$ es $O * O$, por lo que $P + (-P) = (P * (-P)) * O = (O * O) * O = O + O = 0$, donde la última igualdad sigue de que O es el elemento neutro de $+$.

Comprobemos ahora la asociatividad, esto es, que dados $P, Q, R \in E$ arbitrarios, se tiene $(P + Q) + R = P + (Q + R)$. En primer lugar, nótese que $(P + Q) + R = ((P + Q) * R) * O$ y que $P + (Q + R) = (P * (Q + R)) * O$, por lo que basta ver que $(P + Q) * R = P * (Q + R)$.

Consideremos las rectas que unen P, Q y $P * Q$; $P + Q$, R y $(P + Q) * R$; y $Q * R$, O y $Q + R$. Dado que cada una de ellas está determinada por una ecuación lineal, su producto nos da una ecuación cúbica cuyo conjunto de soluciones, al que llamaremos C_1, es la curva formada por la unión de las tres rectas. Lo mismo ocurre para las rectas que unen $P * Q$, O y $P + Q$; Q, R y $Q * R$; y P, $Q + R$ y $P * (Q + R)$, cuya unión denotaremos por C_2.

Por construcción, por cada uno de los puntos O, P, Q, R, $P * Q$, $Q * R$, $P + Q$ y $Q + R$ pasa una de las rectas de C_1 y una de las de C_2 (obsérvese la Figura 3.2), lo cual implica estos son ocho de los nueve puntos de intersección de C_1 y C_2 que nos da el teorema de Bézout. Ahora bien, E también pasa por estos ocho puntos, de modo que, por el lema precedente, debe pasar por el noveno punto, que solo puede ser la intersección de la recta que pasa por P, $Q + R$ y $P * (Q + R)$ y la que pasa por $P + Q$, R y $(P + Q) * R$. Como, por el teorema de Bézout, C tiene exactamente nueve puntos en común con C_1 y con C_2 esto fuerza a que la novena intersección sea $P * (Q + R) = (P + Q) * R \in E$.

Por último, la conmutatividad es clara, pues la recta que pasa por P y Q es la misma que la que pasa por Q y P, de donde $P * Q = Q * P$, que implica que $P * Q * O = (Q * P) * O$. Equivalentemente, $P + Q = Q + P$. □

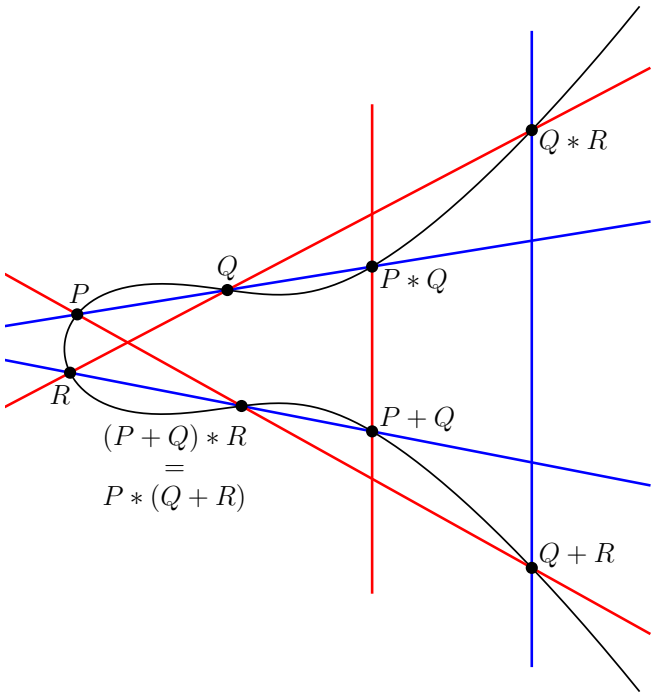


Figura 3.2: Propiedad asociativa de la ley de grupo, tomando O como el punto del infinito.

Observación 3.4. Con la notación de la definición 3.1 y como ya hicimos en el capítulo 2, es habitual tomar O como el punto del infinito $(0 : 1 : 0)$ (y es lo que emplearemos de nuevo durante el capítulo 5). Esto tiene una serie de consecuencias inmediatas:

(i) $O * O = O$, pues O es un punto de inflexión de E, como vimos en el ejemplo 1.47.

(ii) Si E está dado en forma afín por una ecuación de la forma $y^2 = x^3 + ax + b$, $a, b \in K$, entonces el elemento opuesto de un punto afín $P = (x_0, y_0)$ es $-P = (x_0, -y_0)$, es decir, el simétrico respecto al eje x.

No es necesario trabajar sobre un cuerpo algebraicamente cerrado para dar una estructura de grupo a una curva elíptica. En efecto, si E es una curva elíptica definida sobre un subcuerpo $K \subset \overline{K}$ y $O \in E(K)$ (donde recordemos que con $E(K)$ denotamos el conjunto de puntos K-racionales de E), entonces $(E(K), +)$ es un subgrupo de $(E, +)$, pues si $P, Q \in E(K)$, entonces $P + Q \in E(K)$ y $-P \in E(K)$. Esto es debido a que la ecuación de la recta L que conecta dos puntos con coordenadas en K ha de tener coeficientes en K y, si E está definida sobre K, entonces las coordenadas del tercer punto de intersección entre L y E están dadas por una combinación racional de los coeficientes de L y E, y por tanto están en K.

3.2. Endomorfismos

Volviendo al estudio de las aplicaciones entre curvas elípticas y sabiendo ahora que estas poseen una estructura de grupo, es natural preguntarse qué isogenias son homomorfismos. La respuesta es que todas ellas lo son:

Teorema 3.5. *Sea $\phi : E(\overline{K}) \to E'(\overline{K})$ una isogenia entre dos curvas elípticas E y E' sobre K. Entonces $\phi(P+Q) = \phi(P) + \phi(Q)$ para todo $P, Q \in E$.*

Demostración. Véase [8, Theorem III.4.8]. □

En el caso particular en el que $E = E'$, hablaremos de *endomorfismos* de E.

Definición 3.6. Sea E una curva elíptica sobre K. Un *endomorfismo* de E es un morfismo de curvas $\phi : E(\overline{K}) \to E(\overline{K})$ que fija O.

Podemos definir la suma de dos isogenias $\phi_1, \phi_2 : E(\overline{K}) \to E'(\overline{K})$ como la isogenia $\phi_1 + \phi_2 : E(\overline{K}) \to E'(\overline{K})$ dada por $(\phi_1 + \phi_2)(P) = \phi_1(P) + \phi_2(P)$. Cuando $E(\overline{K}) = E'(\overline{K})$, las isogenias forman un anillo, con la suma y composición de isogenias como suma y producto, respectivamente. Este se conoce como el *anillo de endomorfismos* de E y se denota por $\mathrm{End}(E)$.

Ejemplo 3.7. Sea E una curva elíptica sobre K. La aplicación "multiplicación por n", que lleva un punto $P \in E(\overline{K})$ a $n(P) = nP$, donde nP denota la suma de n copias de P, es un endomorfismo de E.[2]

Por la definición de isogenia, dado un endomorfismo ϕ de E, existen funciones racionales R_1, R_2 con coeficientes en $\overline{K}$ tales que $\phi(x,y) = (R_1(x,y), R_2(x,y))$ para los puntos (x,y) afines de $E(\overline{K})$. En el caso de que E esté dada por la forma de Weierstrass $y^2 = x^3 + Ax + B$, podemos dar una forma estándar para cualquier endomorfismo:

Proposición 3.8. *Consideremos una curva elíptica $E : y^2 = x^3 + Ax + B$, con $A, B \in K$, y sea $\phi : E(\overline{K}) \to E(\overline{K})$ un endomorfismo de E.*

Entonces, existen funciones racionales $f_1, f_2 \in \overline{K}(x)$ tales que $\phi(x,y) = (f_1(x), f_2(x)y)$ para los puntos (x,y) afines de $E(\overline{K})$. Además, si f_1 está dada por $f_1(x) = p(x)/q(x)$, con $p, q \in \overline{K}[x]$ sin factores comunes, entonces $f_2(x)$ está definida si $q(x) \neq 0$ y podemos escribir $\phi(x,y) = O$ si $q(x) = 0$.

[2] Por consiguiente, $\mathrm{End}(E)$ contiene un subanillo isomorfo a $\mathbb{Z}$. Si $\mathrm{End}(E)$ es estrictamente mayor que $\mathbb{Z}$, entonces se dice que E tiene *multiplicación compleja*. Las curvas elípticas con multiplicación compleja poseen numerosas propiedades de gran interés [8, Remark III.4.3], [10, Chapter 10].

Demostración. Sea R una función racional en las variables x e y. Puesto que todo punto afín de E satisface la ecuación $y^2 = x^3 + Ax + B$, podemos reemplazar cualquier potencia par de y por un polinomio en x, y cualquier potencia impar de y por y multiplicado por un polinomio en x. Así, la expresión de R se puede reescribir como

$$R(x,y) = \frac{p_1(x) + p_2(x)y}{p_3(x) + p_4(x)y},$$

para oportunos $p_1, p_2, p_3, p_4 \in \overline{K}[x]$. Si se racionaliza multiplicando el numerador y denominador por $p_3 - p_4 y$ y, a continuación, se vuelve a sustituir y^2 por $x^3 + Ax + B$, se obtiene

$$R(x,y) = \frac{q_1(x) + q_2(x)y}{q_3(x)},$$

con oportunos $q_1, q_2, q_3 \in \overline{K}[x]$.

Dado que ϕ es, en particular, un homomorfismo, se tiene que $\phi(x,-y) = \phi(-(x,y)) = -\phi(x,y)$. Si R_1 y R_2 son funciones racionales tales que $\phi(x,y) = (R_1(x,y), R_2(x,y))$, entonces $R_1(x,-y) = R_1(x,y)$ y $R_2(x,-y) = -R_2(x,y)$. Por tanto, R_1 es par en la variable y y R_2 es impar en y, de modo que, al escribirlos en la forma (3.2), se tiene que $q_2 = 0$ para R_1 y $q_1 = 0$ para R_2. Se puede asumir entonces que ϕ está dada por

$$\phi(x,y) = (f_1(x), f_2(x)y),$$

donde $f_1, f_2 \in \overline{K}(x)$ son funciones racionales.

Escribiendo $f_1(x) = p(x)/q(x)$, con $p, q \in \overline{K}[x]$ sin factores comunes, definimos $\phi(x,y) = O$ si $q(x) = 0$. Si $q(x) \neq 0$, entonces se puede demostrar que $f_2(x)$ está definida [10, Section 2.9]. □

Definición 3.9. Sea ϕ un endomorfismo de E. Con la notación de la proposición precedente, definimos el *grado* de ϕ como

$$\deg \phi := \text{máx}\{\deg p, \deg q\}$$

si ϕ no es la isogenia nula. De lo contrario, definimos $\deg [0] = 0$.

Definición 3.10. Sea $\phi \neq [0]$ un endomorfismo de E. Con la notación de la proposición precedente, decimos que ϕ es *separable* si la derivada f_1' no es idénticamente nula.

La siguiente proposición, que nos da una relación entre el grado de un endomorfismo y el número de puntos en su núcleo, nos será de utilidad para probar el teorema de Hasse en el capítulo 4, el cual da una cota del número de puntos de una curva elíptica definida sobre un cuerpo finito.

Proposición 3.11. *Sea E una curva elíptica sobre K y $\alpha \neq [0]$ un endomorfismo de E. Si α es separable, se tiene que* $\deg \alpha = \#\ker \alpha$. *Si no es separable,* $\deg \alpha > \#\ker \alpha$.

Demostración. Por la proposición 3.8, podemos escribir $\alpha(x,y) = (f_1(x), f_2(x)y)$ para ciertos $f_1, f_2 \in \overline{K}(x)$ y para todo punto afín (x,y) de $E(\overline{K})$. Sea $f_1(x) = p(x)/q(x)$, con $p, q \in \overline{K}[x]$. Si α es separable, entonces $f_1' \neq 0$, esto es, $p'q - pq' \neq 0$.

Consideremos el conjunto $S := \{ x \in \overline{K} : (p'q - pq')(x)\, q(x) = 0 \}$. Dado que $p'q - pq' \neq 0$, S es finito, y por tanto $\alpha(S)$ es también finito. En contraposición, $\alpha(E(\overline{K}))$ es infinito, ya que toda isogenia no nula es sobreyectiva. Por tanto, existe un punto $(a : b : 1) \in \alpha(E(\overline{K})) = E(\overline{K})$ tal que

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

(b) $\deg(p(x) - aq(x)) = \deg \alpha$, y

(c) $a \notin f_1(S)$.

Consideremos ahora el conjunto $R := \{ (x_1 : y_1 : 1) \in E(\overline{K}) : \alpha(x_1, y_1) = (a, b) \}$, y veamos que R tiene $\deg \alpha$ elementos.

Dado $(x_1 : y_1 : 1) \in R$, se tiene que

$$\frac{p(x_1)}{q(x_1)} = a, \quad y_1 f_2(x_1) = b.$$

Como $(a : b : 1)$ no es el punto del infinito, $q(x_1) \neq 0$ y, por tanto, por la proposición 3.8, $f_2(x_1)$ está definido. Como además, por (a), $y_1 f_2(x_1) = b \neq 0$, tenemos que $y_1 = b/f_2(x_1)$ y que y_1 está determinado por x_1. Así, $\#R = \#\{ x : \exists y \in \overline{K} \text{ tal que } (x : y : 1) \in R \}$.

Veamos ahora que las raíces de $p(x) - aq(x) = 0$ son todas distintas entre sí. Supongamos que existe una raíz x^* con multiplicidad mayor que uno. Entonces

$$p(x^*) = aq(x^*) \quad \text{y} \quad p'(x^*) = aq'(x^*).$$

Multiplicando ambas ecuaciones, se llega a la expresión $ap(x^*)q'(x^*) = ap'(x^*)q(x^*)$. Por la condición (a), $a \neq 0$, de modo que podemos dividir a ambos lados de la ecuación y llegar a que x^* es raíz del polinomio $pq' - p'q$. Sigue que $f_1(x^*) = a \in f_1(S)$, que contradice (c). Por tanto, usando también (b), $p - aq$ tiene $\deg \alpha$ raíces distintas. Entonces, existen $\deg \alpha$ puntos (x_1, y_1) tales que $\alpha(x_1, y_1) = (a, b)$ y $\#R = \deg \alpha$.

Dado que α es un homomorfismo, este razonamiento es en realidad válido para cualquier punto de $\alpha(E(\overline{K})) = E(\overline{K})$, de modo que el núcleo de α tiene $\deg \alpha$ elementos.

Si α no es separable, $p' - aq' = 0$ y $p - aq$ tiene raíces múltiples. El resto del razonamiento sigue siendo válido, por lo que se concluye que el número de soluciones de $p - aq$ es menor que $\deg \alpha$, y de ahí que $\#\ker \alpha < \deg \alpha$. □

El siguiente lema, cuya demostración es técnica y no incluimos, nos permitirá caracterizar la separabilidad de los endomorfismos multiplicación por n. A su vez, esto resultará clave para demostrar el teorema de Hasse en el capítulo 4.

Lema 3.12. *Sea E una curva elíptica sobre K, y sean α_1, α_2 y α_3 endomorfismos no nulos de E tales que $\alpha_1 + \alpha_2 = \alpha_3$. Sean f_{α_j} y g_{α_j} funciones racionales tales que $\alpha_j(x,y) = (f_{\alpha_j}(x), g_{\alpha_j}(x)y)$ para todo punto afín (x,y) de E, con $j = 1,2,3$.*

Supongamos que existen constantes $c_{\alpha_1}, c_{\alpha_2}$ tales que

$$\frac{f'_{\alpha_1}(x)}{g_{\alpha_1}(x)} = c_{\alpha_1} \quad y \quad \frac{f'_{\alpha_2}(x)}{g_{\alpha_2}(x)} = c_{\alpha_2}.$$

Se tiene que

$$\frac{f'_{\alpha_3}(x)}{g_{\alpha_3}(x)} = c_{\alpha_1} + c_{\alpha_2}.$$

Demostración. Se puede encontrar en [10, Lemma 2.26]. □

Proposición 3.13. *Sean E una curva elíptica sobre K y n un entero no nulo. Supongamos que el endomorfismo multiplicación por n en E está dado por*

$$n(x,y) = (f_n(x), g_n(x)y)$$

para todo punto afín (x,y) de $E(\overline{K})$, con f_n, g_n funciones racionales. Entonces, se tiene que

$$\frac{f'_n(x)}{g_n(x)} = n. \tag{3.1}$$

Por consiguiente, la multiplicación por n es separable si y solo si n no es múltiplo de la característica de K.

Demostración. En primer lugar, nótese que basta probar el resultado para $n > 0$, pues, por el mismo razonamiento que en la demostración de la proposición 3.8, $f_{-n} = f_n$ y $g_{-n} = -g_n$, lo cual implica que $f'_{-n}/g_{-n} = -f'_n/g_n$.

La ecuación (3.1) es trivialmente cierta para $n = 1$. Además, si suponemos que se cumple para n, entonces es cierta para $n+1$ por el lema 3.12, pues la multiplicación por $n+1$ es la suma de las multiplicaciones por n y 1. Así, la primera parte de la proposición sigue por inducción sobre n.

Con respecto a la segunda parte, basta notar que la multiplicación por n es separable si y solo si $f'_n \neq 0$, y que, por la ecuación (3.1), esto es cierto si y solo si $n = f'_n/g_n \neq 0$, que es equivalente a que n no sea divisible por la característica de K. □

3.3. Puntos racionales y enteros

El estudio de las soluciones racionales y enteras de curvas tiene una larga tradición, cuyos orígenes se remontan a la geometría diofántica, a la que da nombre el matemático griego Diofanto de

Alejandría. En 1890, Hilbert y Hurwitz demostraron que si una curva de género cero tiene un punto $\mathbb{Q}$-racional, más comúnmente referido como *racional*, entonces tiene infinitos puntos racionales, y que estos están dados por los valores racionales de un parámetro. En 1901, Poincaré publica un largo artículo sobre puntos racionales en curvas e introduce la cuestión de si estos son finitamente generados. Beppo Levi fue el primero en plantear si esta afirmación, que Poincaré daba por cierta, era verdadera.

En 1922, Mordell responde afirmativamente a la pregunta de Beppo Levi con uno de los teoremas más importantes de la geometría diofántica, conocido como "teorema de Mordell" (generalizado a cualquier cuerpo de números por Weil). Da nombre también a la "conjetura de Mordell", que afirma que todas las curvas de género mayor que uno sobre $\mathbb{Q}$ tienen un número finito de puntos racionales. Dicha conjetura fue demostrada por Faltings en 1983, y es uno de los resultados más importantes de la geometría algebraica aritmética moderna.

Los resultados presentados en la presente sección tienen demostraciones complejas que escapan a los objetivos de este trabajo. Nos limitamos a enunciarlos y a dar un breve comentario sobre su significado.

Teorema 3.14 (Teorema de Mordell-Weil)**.** *Sea K un cuerpo de números, es decir, una extensión finita de $\mathbb{Q}$, y sea E una curva elíptica sobre K. Entonces, el grupo $E(K)$ está finitamente generado.*

Demostración. Se puede consultar en [7, Chapter IV] o [8, Chapter VIII]. □

Como consecuencia inmediata, se tiene que

$$E(K) \cong \mathbb{Z}^r \oplus E(K)_{\text{tors}},$$

donde el *rango* r (sobre K) es un entero no negativo y el *subgrupo de torsión* $E(K)_{\text{tors}}$ es finito.

Cuando $K = \mathbb{Q}$, es relativamente sencillo calcular $E(\mathbb{Q})_{\text{tors}}$, gracias al teorema de Lutz-Nagell, y un profundo teorema de Mazur indica cuáles son los grupos que pueden aparecer como subgrupos de torsión de curvas elípticas. En contraposición, se desconoce un método general para calcular el rango r, y se cree que puede ser arbitrariamente grande [20]. La curva con el mayor rango conocido, de al menos 29, fue encontrada por Elkies y Klagsburn en 2024 [14, 13]. El rango es también uno de los objetos de interés de uno de los problemas del milenio, la conjetura de Birch y Swinnerton-Dyer.

Observación 3.15. Nótese que el número de puntos racionales de una curva elíptica es infinito si y solo si su rango sobre $\mathbb{Q}$ es positivo.

Continuando con el estudio de las curvas elípticas desde el punto de vista de la geometría diofántica, una pregunta esperable es qué se puede decir sobre los puntos *enteros* (es decir, $\mathbb{Z}$-racionales) de su parte afín. Esto es relevante, entre otros motivos, porque una de las condiciones necesarias para

que un punto sea de orden finito (*de torsión*) es que sus coordenadas sean enteras, según el teorema de Luzt-Nagell. El teorema de Siegel de 1929 da un importante resultado al respecto:

Teorema 3.16 (Teorema de Siegel)**.** *Sea E una curva elíptica cuya parte afín viene dada por $C : y^2 = x^3 + Ax + B$, con A y B enteros. Entonces, C tiene un número finito de puntos enteros.*

Demostración. Véase [31]. □

Observación 3.17. En el teorema de Siegel es importante considerar la parte afín de la curva elíptica. Por ejemplo, si E es una curva elíptica sobre $\mathbb{Q}$ de rango $r > 0$, entonces, por la observación 3.15, E tiene infinitos puntos racionales y, por tanto, considerando coordenadas proyectivas y eliminando denominadores, también infinitos puntos enteros.

Aunque hemos presentado el teorema de Siegel en su versión para curvas elípticas, el propio Siegel dio una versión más general, que dice que una curva (afín) no singular definida sobre un cuerpo de números y de género $g > 0$ tiene un número finito de puntos enteros [19, Theorem D.9.1]. Para el caso $g \geq 2$, el teorema de Faltings generaliza a su vez este resultado:

Teorema 3.18 (Teorema de Faltings)**.** *Sea C una curva no singular de grado estrictamente mayor que 3 (equivalentemente, de género $g \geq 2$) sobre $\mathbb{Q}$. Entonces $C(\mathbb{Q})$ es finito.*

Demostración. Puede consultarse en [19, Part E]. □

3.4. Puntos de torsión

A modo de paréntesis, incluimos a continuación un estudio sobre los puntos de torsión de una curva elíptica E definida sobre un cuerpo K arbitrario, esto es, los puntos P de E tales que $nP = O$ para un cierto entero $n > 0$. Este análisis no será empleado en el resto del trabajo, pero se incluye tanto por la relevancia de los resultados que se presentan como por constituir un buen ejemplo del uso de herramientas propias de la teoría de números en el estudio de las curvas elípticas.

La mayor parte de esta sección estará dedicada a probar el teorema de Lutz-Nagell, que proporciona propiedades importantes sobre estos puntos. Comentaremos sus principales aplicaciones y finalizaremos enunciando el teorema de Mazur, que caracteriza los subgrupos de torsión de las curvas elípticas.

En primer lugar, analizaremos los puntos de orden 2, que necesitaremos caracterizar para la posterior demostración del teorema de Lutz-Nagell. Empleando la notación habitual de la literatura, estaremos entonces analizando el caso $n = 2$ de los siguientes conjuntos:

$$E[n] = \{ P \in E(\overline{K}) : nP = O \},$$

definidos para $n \in \mathbb{Z}$, $n > 0$. Nótese que $E[n]$ es un subgrupo de $E(\overline{K})$, al tratarse del núcleo del endomorfismo multiplicación por n.

Proposición 3.19. *Sea E una curva elíptica sobre K. Si char $K \neq 2$, entonces $E[2] \cong \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$, y se tiene que dado $P = (x : y : 1) \in E(\overline{K})$, $P \in E[2]$ si y solo si $y = 0$.*

Demostración. Si suponemos char $K \neq 2$, podemos escribir la forma afín de $E(\overline{K})$ con la ecuación $y^2 = (x-\alpha)(x-\beta)(x-\gamma)$, como hicimos en la demostración de la proposición 2.15. Dado $P = (x : y : 1) \in E(\overline{K})$, tenemos que la condición $2P = O$ es equivalente a que $P = -P$. Como $-P = (x : -y : 1)$, esto ocurre si y solo si $y = 0$, de forma que $E[2] = \{O, (\alpha : 0 : 1), (\beta : 0 : 1), (\gamma : 0 : 1)\}$, esto es, $E[2]$ es el subgrupo formado por el punto del infinito y los puntos de E con coordenada y nula.

Como los elementos de $E[2]$ tienen todos orden 1 o 2, se deduce que $E[2] \cong \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$. □

Observación 3.20. El conjunto de puntos de orden 2 con coordenadas en K en lugar de en $\overline{K}$ no necesariamente tiene estructura isomorfa a $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$, pero sí es cierto que $P = (x : y : 1) \in E(K)$ tiene orden 2 si y solo si $y = 0$.

La anterior proposición tiene la siguiente generalización, que emplearemos en el capítulo 4 para estudiar la estructura del grupo de una curva elíptica sobre un cuerpo finito:

Teorema 3.21. *Sea E una curva elíptica sobre K y sea n un entero positivo.*

Si char $K \nmid n$ o char $K = 0$, entonces $E[n] \cong \mathbb{Z}/n\mathbb{Z} \oplus \mathbb{Z}/n\mathbb{Z}$.

Si $p := \text{char } K > 0$ y $p \mid n$, escribiendo $n = p^r n'$ con $p \nmid n'$, entonces $E[n] \cong \mathbb{Z}/n'\mathbb{Z} \oplus \mathbb{Z}/n'\mathbb{Z}$ o $E[n] \cong \mathbb{Z}/n\mathbb{Z} \oplus \mathbb{Z}/n'\mathbb{Z}$.

Demostración. Por su complejidad, remitimos esta demostración a [10, Section 3.2]. □

Como veremos, la primera parte del teorema de Lutz-Nagell enuncia que los puntos de torsión racionales de una curva elíptica tienen coordenadas enteras. Para demostrarla, argumentaremos que los denominadores de dichas coordenadas no son divisibles por ningún número primo, lo cual nos permitirá concluir que son enteras. Con este objetivo, introducimos a continuación conceptos p-ádicos que nos serán de utilidad a la hora de estudiar la divisibilidad de números por primos. Nos limitaremos a dar una visión superficial de los mismos, pues las numerosas conexiones entre la teoría de números p-ádicos y las curvas elípticas no son el punto central de este trabajo.

Sea p un primo y $a \in \mathbb{Q}$, $a \neq 0$. Entonces a se puede escribir de forma única como $a = p^r \frac{m}{n}$, con $m, n \in \mathbb{Z}$, coprimos y no divisibles por p. La *valoración p-ádica* de a se define como

$$v_p(a) = r,$$

y definimos también $v_p(0) = +\infty$.

Sea E una curva elíptica dada por $Y^2Z = X^3 + AXZ^2 + BZ^3$, con $A, B \in \mathbb{Z}$, y sea $(x : y : 1) \in E$ un punto racional. Veamos que, si p divide el denominador de x, entonces divide al de y, y viceversa. Supongamos que x e y están dados por

$$x = \frac{m}{np^\mu} \quad \text{e} \quad y = \frac{u}{wp^\sigma},$$

con $\mu > 0$ y $p \nmid m, n, u, w$. Sustituyendo en la ecuación de E, se obtiene que

$$\frac{u^2}{w^2p^{2\sigma}} = \frac{m^3 + am^2np^\mu + bmn^2p^{2\mu} + cn^3p^{3\mu}}{n^3p^{3\mu}}.$$

Que $p \nmid u^2, w^2$ implica que $v_p(\frac{u^2}{w^2p^{2\sigma}}) = -2\sigma$. Y del hecho de que $\mu > 0$ y $p \nmid m$ se deduce que $p \nmid m^3 + am^2np^\mu + bmn^2p^{2\mu} + cn^3p^{3\mu}$. Por tanto, $v_p(\frac{m^3+am^2np^\mu+bmn^2p^{2\mu}+cn^3p^{3\mu}}{n^3p^{3\mu}}) = -3\mu$.

Por consiguiente, $2\sigma = 3\mu$, de donde $\sigma > 0$, $2 \mid \mu$ y $3 \mid \sigma$, de modo que $\mu = 2n$ y $\sigma = 3n$ para un cierto entero $n > 0$.

Suponiendo que p divide el denominador de y, con un razonamiento análogo, se deduce que si p divide el denominador de x o de y, entonces divide el denominador de ambos y que $v_p(x) = -2n$ y $v_p(y) = -3n$ para un cierto entero $n > 0$.

Este resultado motiva la siguiente definición:

$$E^n(\mathbb{Q}) = \{(x : y : 1) \in E(\mathbb{Q}) : v_p(x) \leq -2n, v_p(y) \leq -3n\} \cup \{O\}.$$

Los conjuntos $E^n(\mathbb{Q})$ contienen puntos cercanos a O módulo potencias de p (es decir, que son p-ádicamente cercanos a O).

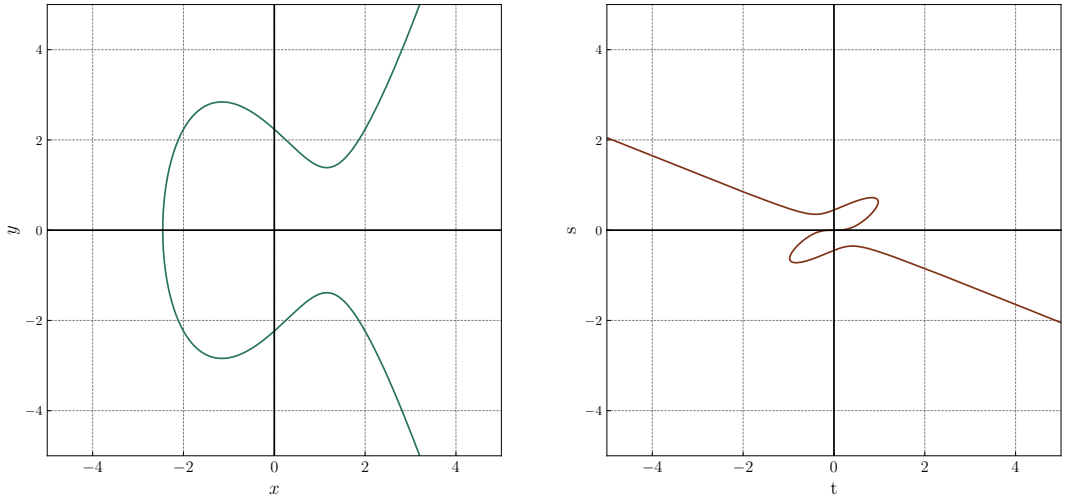
De la definición es clara la siguiente cadena de inclusiones (que es un ejemplo de lo que en la literatura se conoce como “filtración p-ádica”):

$$E(\mathbb{Q}) \supset E^1(\mathbb{Q}) \supset E^2(\mathbb{Q}) \supset \ldots \text{ con } \bigcap_{n=1}^{\infty} E^n(\mathbb{Q}) = O.$$

En lo que sigue, denotaremos por R el anillo de elementos p-integrales de $\mathbb{Q}$, es decir, los números racionales tales que p no divide sus denominadores.

Sea $P = (x : y : 1)$ un punto racional. Consideremos con el cambio de coordenadas

$$t(P) = \frac{x}{y}, \quad s(P) = \frac{1}{y} \tag{3.2}$$



(a) Con coordenadas (x, y).

(b) Con coordenadas (t, s).

Figura 3.3: Curva elíptica definida sobre $\mathbb{R}$, dada por $y^2 = x^3 - 4x + 5$ y por $s = t^3 - 4ts^2 + 5s^3$ tras aplicar el cambio de coordenadas 3.2.

y supongamos que $P \in E^n(\mathbb{Q})$. Por definición de $E^n(\mathbb{Q})$, x e y son de la forma $x = mn^{-1}p^{-2(n+i)}$ e $y = uw^{-1}p^{-3(n+i)}$, con $p \mid m, n, u, w$ e $i \geq 0$. Entonces,

$$t(P) = \frac{mw}{nu}p^{n+i} \quad \text{y} \quad s(P) = \frac{w}{u}p^{3(n+i)}. \tag{3.3}$$

Por consiguiente, $P \in E^n(\mathbb{Q})$ si y solo si $t(P) \in p^n R$ y $s(P) \in p^{3n}R$.

Además, definimos la imagen de O por el cambio de coordenadas como $(0, 0)$. La Figura 3.3 muestra un ejemplo del resultado de aplicar el cambio a una curva elíptica.

Calculando fórmulas explícitas para la suma de dos puntos en coordenadas (t, s), se demuestra el siguiente lema:

Lema 3.22. *Sean $P_1 = (x_1 : y_1 : 1), P_2 = (x_2 : y_2 : 1) \in E^n(\mathbb{Q})$. Entonces, $t(P_1) + t(P_2) + t(P_1 * P_2) \in p^{3n}R$.*

Demostración. Puede consultarse en [9, Section 2.4]. □

Proposición 3.23. *Sea p un primo. Se tiene que:*

1. *Para todo $n \geq 1$, $E^n(\mathbb{Q})$ es un subgrupo de $E(\mathbb{Q})$.*
2. *La aplicación*

$$\begin{aligned} \frac{E^n(\mathbb{Q})}{E^{3n}(\mathbb{Q})} &\longrightarrow \frac{p^n R}{p^{3n}R} \\ P = (x : y : 1) &\longmapsto t(P) = \frac{x}{y} \\ O = (0 : 1 : 0) &\longmapsto t(O) = 0 \end{aligned}$$

es un homomorfismo inyectivo.

Demostración. 1. Sean $P_1 = (x_1 : y_1 : 1)$, $P_2 = (x_2 : y_2 : 1) \in E^n(\mathbb{Q})$. Por el lema anterior, puesto que $t(P_1), t(P_2) \in p^n R$, sigue que $t(P_1 * P_2) \in p^n R$. Pero entonces, en vista de las fórmulas de (3.3), $s(P_1 * P_2) \in p^{3n}R$ y, por tanto, $P_1 * P_2 \in E^n(\mathbb{Q})$. Escribiendo $P_1 * P_2 = (x_3 : y_3 : 1)$, por el punto (ii) de la observación 3.4 se tiene que $P_1 + P_2 = (x_3 : -y_3 : 1) \in E^n(\mathbb{Q})$.

Además, $P_1 + O = (x_1 : -y_1 : 1) \in E^n(\mathbb{Q})$ y $O + O = O \in E^n(\mathbb{Q})$.

Concluimos que $E^n(\mathbb{Q})$ es un subgrupo de $E(\mathbb{Q})$.

2. En primer lugar, obsérvese que $t(P_1 + P_2) = -t(P_1 * P_2)$. Por tanto, sigue del lema anterior que $t(P_1) + t(P_2) - t(P_1 + P_2) \in p^{3n}R$. Equivalentemente, $t(P_1 + P_2) \equiv t(P_1) + t(P_2) \pmod{p^{3n}R}$. Esto es cierto también en el caso en el que $P_1 = O$ o $P_2 = O$, de donde la composición $P \longmapsto t(P) \longmapsto t(P) + p^{3n}R$ es un homomorfismo.

Concluimos la prueba notando que el núcleo de este homomorfismo es el conjunto de puntos P tales que $t(P) \in p^{3n}R$, es decir, $E^{3n}(\mathbb{Q})$. □

Corolario 3.24. *Para todo p primo,* $E^1(\mathbb{Q}) \cap E(\mathbb{Q})_{tors} = \{O\}$.

Demostración. Fijemos un primo p. Si $E^1(\mathbb{Q}) \cap E(\mathbb{Q})_{\text{tors}} \neq \{O\}$, entonces existe un punto $P \neq O$ de orden finito q tal que $P \in E^1(\mathbb{Q})$.

Sabemos que $\bigcap_{n=1}^{\infty} E^n(\mathbb{Q}) = O$, por lo que existe $n > 0$ tal que $P \in E^n(\mathbb{Q})$, pero $P \notin E^{n+1}(\mathbb{Q})$.

Supongamos que $p \nmid q$. Puesto que $qP = O$ y por ser la aplicación de la parte (ii) de la proposición 3.23 un homomorfismo, se tiene que $t(qP) = t(O) = 0$.

Por otro lado, aplicando repetidas veces la congruencia $t(P_1 + P_2) \equiv t(P_1) + t(P_2) \pmod{p^{3n}R}$, obtenemos que $0 = t(qP) \equiv qt(P) \pmod{p^{3n}R}$. Como $p \nmid q$, q es una unidad de R y deducimos que $t(P) \equiv 0 \pmod{p^{3n}R}$.

Pero entonces $P \in E^{3n}(\mathbb{Q})$, que contradice $P \notin E^{n+1}(\mathbb{Q})$.

Supongamos ahora que $q = pr$ para un entero $r \geq 1$, y consideremos el punto $P' = rP$. Como P tiene orden q, P' ha de tener orden p. Además, $P \in E^1(\mathbb{Q})$, que es un subgrupo de $E(\mathbb{Q})$, lo cual implica que $P' \in E^1(\mathbb{Q})$. Razonando como antes, existe $n' > 0$ tal que $P' \in E^{n'}(\mathbb{Q})$, pero $P' \notin E^{n'+1}(\mathbb{Q})$. Llegamos entonces a que $0 = t(O) = t(pP') \equiv pt(P') \pmod{p^{3n'}R}$ y, de esto, $t(P') \equiv 0 \pmod{p^{3n'-1}R}$.

Pero $3n' - 1 \geq n' + 1$, que contradice $P' \notin E^{n'+1}(\mathbb{Q})$. □

Estamos en condiciones de demostrar el siguiente teorema, probado de forma independiente por Lutz y Nagell en la década de 1930:

Teorema 3.25 (Teorema de Lutz-Nagell). *Sea E una curva elíptica dada por la ecuación $y^2 = x^3 + Ax + B$, con A y B enteros. Si $P = (x_1 : y_1 : 1) \in E(\mathbb{Q})_{(tors)}$, entonces*

(i) $x_1, y_1 \in \mathbb{Z}$, *y*

(ii) $y_1 = 0$ *o* $y_1^2 \mid \Delta$

Demostración. (i) Sea P como en el enunciado. Por el corolario anterior, se tiene $P \notin E^1(\mathbb{Q})$ para el subgrupo $E^1(\mathbb{Q})$ asociado a cualquier primo. Es decir, los denominadores de x_1 e y_1 no son divisibles por ningún primo. Por consiguiente, $x_1, y_1 \in \mathbb{Z}$.

(ii) Supongamos que $y_1 \neq 0$, que implica que $2P \neq O$ (por la observación 3.20). Usando (i) sobre P y $2P := (x_2 : y_2 : 1)$, obtenemos que $x_2, y_2 \in \mathbb{Z}$.

Calculemos ahora una fórmula explícita para x_2 en función de x_1 e y_1. En primer lugar, necesitamos obtener la pendiente λ de la recta tangente a E en P. Derivando implícitamente la ecuación $y^2 = x^3 + Ax + B$, sigue que

$$\lambda = \frac{dy}{dx}\Big|_P = \frac{3x_1^2 + A}{2y_1}.$$

Por tanto, la ecuación de la recta tangente a E en P es $y = \lambda x + \nu$, con $\nu = y_1 - \lambda x_1$. Sustituyendo en la ecuación de E, se tiene que

$$0 = (\lambda x + \nu)^2 - x^3 - Ax - B = -x^3 + \lambda^2 x^2 + (2\lambda\nu - A)x + (\nu^2 - B).$$

Dado que los puntos de intersección de la recta y E son P (con multiplicidad doble) y $2P$, se tiene que x_1, x_1 y x_2 son las tres raíces de la ecuación cúbica anterior. Por tanto, $2x_1 + x_2 = \lambda^2$ y se llega a la siguiente expresión para x_2:

$$x_2 = \lambda^2 - 2x_1 = \frac{(3x_1^2 + A)^2}{4y_1^2} - 2x_1 = \frac{(3x_1^2 + A)^2 - 8x_1y_1^2}{4y_1^2} = \frac{(3x_1^2 + A)^2 - 8x_1(x_1^3 + Ax_1 + B)}{4y_1^2}$$
$$= \frac{x_1^4 - 2Ax_1^2 - 8Bx_1 + A^2}{4y_1^2}.$$

Del hecho de que $x_2 \in \mathbb{Z}$ se deduce que $y_1^2 \mid x_1^4 - 2Ax_1^2 - 8Bx_1 + A^2$.

Por otro lado, mediante sencillas manipulaciones aritméticas se puede demostrar la siguiente igualdad:

$$(3x_1^2 + 4A)(x_1^4 - 2Ax_1^2 - 8Bx_1 + A^2) - (3x_1^3 - 5Ax_1 - 27B)(x_1^3 + Ax_1 + B) = 4A^3 + 27B^2.$$

Puesto que $y_1^2 = x_1^3 + Ax_1 + B$, y_1^2 divide los dos términos del lado izquierdo de la igualdad anterior. Concluimos que $y_1^2 \mid 4A^3 + 27B^2$. □

Observación 3.26. (a) El recíproco del teorema de Lutz-Nagell no es cierto: un punto $P \in E(\mathbb{Q})$ puede satisfacer las condiciones (i) y (ii) del teorema y no ser de torsión.

(b) A menudo, este teorema permite probar que un punto $P \in E(\mathbb{Q})$ tiene orden infinito. Si se calculan los múltiplos nP de P y se encuentra uno tal que sus coordenas x o y no son enteras, entonces P no es de torsión.

Adicionalmente, el teorema de Lutz-Nagell proporciona un método para encontrar todos los puntos de torsión de una curva elíptica dada: para $y = 0$ y para cada y tal que $y^2 \mid \Delta$, se calculan las soluciones x enteras de la ecuación $x^3 + Ax + B = y^2$, y a continuación se comprueba si el orden del punto $(x : y : 1)$ es finito. El siguiente ejemplo ilustra este método:

Ejemplo 3.27. Sea E la curva elíptica dada por la ecuación $y^2 = x^3 + 4$, cuyo discriminante vale $\Delta = 4A^3 + 27B^2 = 432$. Sea $P = (x : y : 1) \in E(\mathbb{Q})$ un punto de torsión.

Dado que $0 = x^3 + 4$ no tiene soluciones racionales, $y \neq 0$. Por tanto, por el teorema de Lutz-Nagell, si P es un punto de torsión, entonces $y^2 \mid 432$. Las únicas posibilidades son $y = \pm 1, \pm 2, \pm 3, \pm 4, \pm 6, \pm 12$, pero solo para $y = \pm 2$ se obtiene un valor racional para x al sustuir en la ecuación de E.

Por tanto, los únicos posibles puntos de torsión son $(0 : 2 : 1)$ y $(0 : -2 : 1)$. Se puede comprobar que $3(0 : \pm 2 : 1) = O$, de modo que ambos son puntos de torsión de orden 3 y $E(\mathbb{Q})_{\text{tors}}$ es isomorfo al grupo cíclico $\mathbb{Z}/3\mathbb{Z}$.

En 1975, Mazur proporciona una caracterización de los posibles grupos de torsión de una curva elíptica definida sobre $\mathbb{Q}$. No obstante, por la complejidad y dificultad técnica de su demostración, en este trabajo incluimos únicamente el enunciado del resultado.

Teorema 3.28 (Teorema de Mazur)**.** *Sea E una curva elíptica sobre $\mathbb{Q}$. Su subgrupo de torsión $E(\mathbb{Q})_{tors}$ es isomorfo a uno de los siguientes quince grupos:*

$$\begin{aligned} \mathbb{Z}/N\mathbb{Z} \qquad & \text{con } 1 \leq N \leq 10 \text{ o } N = 12, \\ \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2N\mathbb{Z} \quad & \text{con } 1 \leq N \leq 4. \end{aligned}$$

Demostración. Se puede consultar en [24]. □

Observación 3.29. Para cada uno de los grupos indicados en el teorema de Mazur existen infinitas curvas elípticas (con invariantes j distintos) cuyo subgrupo de torsión es isomorfo a tal grupo.

Capítulo 4

Curvas elípticas sobre cuerpos finitos

El contenido de este capítulo está dedicado a las curvas elípticas definidas sobre cuerpos finitos, que son el objeto de estudio de la criptografía con curvas elípticas (a la que dedicaremos el capítulo 5). En concreto, veremos algunas propiedades particulares de su grupo y demostraremos el teorema de Hasse, que da una cota para el número de puntos de la curva. Esta cota tiene importantes aplicaciones a la hora de elegir diversos parámetros en la criptografía de curvas elíptica.

A lo largo de este capítulo, emplearemos la siguiente notación: p será un número primo, $\mathbb{F}_q$ será un cuerpo finito de $q = p^n$ elementos para un entero $n \geq 1$ y $\overline{\mathbb{F}}_q$ será una clausura algebraica fija de $\mathbb{F}_q$.

La principal referencia es el libro de Washington [10], muy enfocado a las curvas elípticas en cuerpos finitos y a sus aplicaciones en criptografía. De forma complementaria, pueden consultarse el libro de Silverman [8] y el de Husemöller [20].

4.1. Estructura de grupo de las curvas elípticas sobre cuerpos finitos

Una curva elíptica sobre $\mathbb{F}_q$ se puede visualizar como un conjuntos de puntos $(x, y) \in \mathbb{F}_q \times \mathbb{F}_q$ al que se le añade un punto del infinito. De manera análoga a como ocurría al considerar curvas elípticas sobre $\mathbb{R}$, hay presente una simetría con respecto a $y = 0$, que es aparente en el ejemplo de la Figura 4.1 si se extiende $\mathbb{F}_q \times \mathbb{F}_q$, teselando el plano.

La interpretación geométrica de la ley de grupo presentada en el capítulo 3 también es posible en este caso. En la Figura 4.2a representamos un ejemplo, tomando como elemento neutro el punto del infinito (de modo que el punto opuesto a un punto dado es el simétrico respecto a $y = 0$). Además, como estamos trabajando sobre $\mathbb{F}_q \times \mathbb{F}_q$, podemos considerar un cuadrado $[0, q] \times [0, q]$ con los extremos opuestos identificados y visualizar la curva sobre un toro. En la Figura 4.2b se muestra el

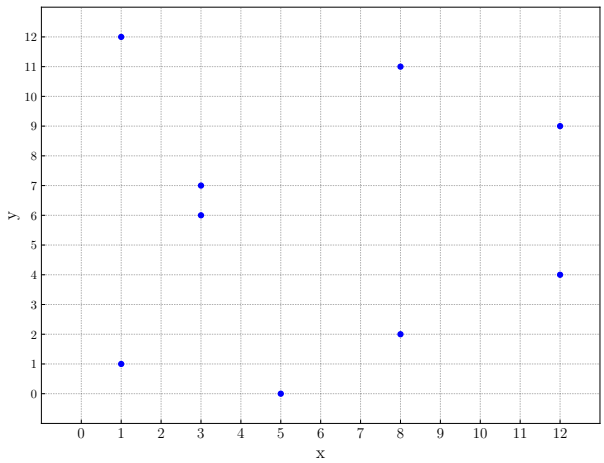


Figura 4.1: Puntos de la curva elíptica $y^2 = x^3 - 2x + 2$ definida sobre $\mathbb{F}_{13}$, excluyendo el punto del infinito.

resultado de realizar esta identificación sobre la curva de la Figura 4.2a.

Por otro lado, el hecho de que el grupo de las curvas elípticas sobre cuerpos finitos sea, a su vez, finito nos permite caracterizar su estructura con más precisión que en el caso general:

Teorema 4.1. *Sea E una curva elíptica sobre $\mathbb{F}_q$. Se tiene que*

$$E(\mathbb{F}_q) \cong \mathbb{Z}/n\mathbb{Z} \text{ o } E(\mathbb{F}_q) \cong \mathbb{Z}/n_1\mathbb{Z} \oplus \mathbb{Z}/n_2\mathbb{Z},$$

para un cierto entero $n \geq 1$ o enteros $n_1, n_2 \geq 1$ con $n_1 | n_2$.

Demostración. Por el teorema de estructura para grupos abelianos finitos, existe un isomorfismo de grupos de la forma

$$E(\mathbb{F}_q) \cong \mathbb{Z}/n_1\mathbb{Z} \oplus \mathbb{Z}/n_2\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/n_r\mathbb{Z}, \tag{4.1}$$

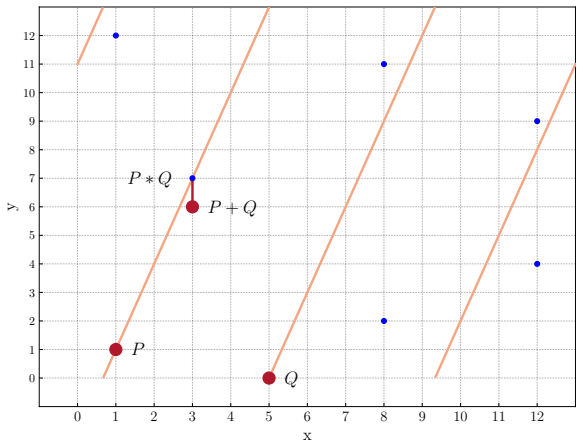
con $n_i \mid n_{i+1}$ para $i = 1, \ldots, r-1$.

Del teorema 3.21 deducimos que, dado un entero positivo n, el subgrupo $E[n]$ tiene, a lo sumo, n^2 elementos. Por consiguiente, $r \leq 2$, pues si $n_3 > 1$, por la ecuación (4.1) tendríamos que $\#E[n_3] \geq n_3^3$, que es una contradicción.

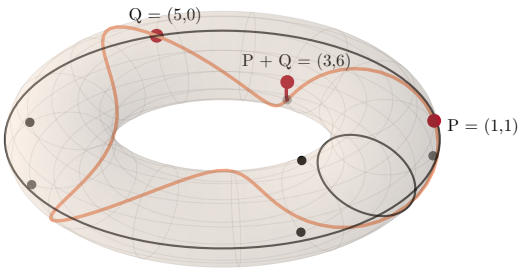
Obsérvese que, si $r = 0$, el grupo es trivial. Esta situación se corresponde con el caso $n = 1$ del enunciado. □

Un concepto relacionado de gran importancia es el de *curvas elípticas supersingulares*:

Definición 4.2. Sea K un cuerpo de característica p (por ejemplo, $K = \mathbb{F}_q$) y E una curva elíptica sobre K. Se dice que E es *supersingular* si $E[p] = \{O\}$. En caso contrario, se dice que E es *ordinaria*.



(a) Suma en $\mathbb{F}_{13}$ representada sobre $[0,13] \times [0,13]$, identificando $(0,y) \sim (13,y)$, para todo $y \in [0,13]$, y $(x,0) \sim (x,13)$, para todo $x \in [0,13]$.



(b) Operación de suma de (a) representada sobre el toro homeomorfo al cuadrado $[0,13] \times [0,13]$ con los ejes opuestos identificados entre sí.

Figura 4.2: Suma de los puntos $P = (1,1)$ y $Q = (5,0)$ de la curva $y^2 = x^3 - 2x + 2$ definida sobre $\mathbb{F}_{13}$, con resultado $P + Q = (3,6)$.

Existen diversas condiciones equivalentes a la de la anterior definición, como que $\mathrm{End}(E)$ sea un álgebra de cuaterniones.

Observación 4.3. No deben confundirse las nociones de singularidad y supersingularidad. Recordamos que, por definición, una curva elíptica es no singular, por lo que una curva elíptica supersingular no puede ser singular. El origen de esta terminología está en que, históricamente, se utilizaba el término "singular" para describir a los invariantes j de aquellas curvas elípticas cuyo anillo de endomorfismos es mayor que $\mathbb{Z}$. Dichos anillos son, habitualmente, subanillos de extensiones cuadráticas de $\mathbb{Q}$. Cuando son subanillos de álgebras de cuaterniones, que son todavía mayores, se empleaba el término "supersingular", con el significado de "muy inusual".

Cuando se trabaja sobre $\mathbb{F}_q$, se tiene la siguiente caracterización para curvas elípticas supersingulares, que se basa en su número de puntos:

Teorema 4.4. *Sea E una curva elíptica sobre $\mathbb{F}_q$. Sea $a = \#E(\mathbb{F}_q) - q - 1$. Entonces, E es supersingular si y solo si $a \equiv 0 \pmod p$, que es equivalente a que $\#E(\mathbb{F}_q) \equiv 1 \pmod p$.*

Demostración. Puede verse en [10, Proposition 4.31]. □

Las curvas elípticas supersingulares son de especial interés en criptografía, tanto por su potencial uso en el diseño de sistemas criptográficos como por su vulnerabilidad frente a algunos ataques [15]. Hacemos énfasis a este respecto en la sección 5.3.

4.2. El teorema de Hasse

Debido a la mayor facilidad de representación de números enteros frente a números en punto flotante en un ordenador, por la mayor eficiencia de las operaciones con los primeros y por los errores de aproximación al trabajar con los segundos, para las aplicaciones en criptografía se trabaja con curvas elípticas sobre cuerpos finitos. Las coordenadas de sus puntos se representan como enteros y las operaciones se realizan aplicando el módulo apropiado.

Por estas aplicaciones criptográficas, es importante saber determinar el número de elementos de una curva elíptica sobre un cuerpo finito $\mathbb{F}_q$. Dado que la ecuación

$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, \quad \text{con } (x,y) \in \mathbb{F}_q^2 \text{ y } a_i \in \mathbb{F}_q$$

tiene, para cada x fijado, a lo sumo dos soluciones, una primera aproximación es que $\#E(F_q) \leq 2q+1$, teniendo en cuenta el punto extra del infinito O.

Más precisamente, tendremos una solución si el discriminante del polinomio cuadrático en la variable y es 0. Si no lo es y además es un cuadrado, tendremos dos soluciones (si ni es nulo ni es un cuadrado, no hay ninguna solución). Teniendo en cuenta que el número de cuadrados en $\mathbb{F}_q$ es $q-1$ si q es par y $(q-1)/2$ si es impar, esperamos que, si la ecuación cuadrática es "aleatoria", el número de puntos es E sea aproximadamente $2(q-1)/2 + 1 + 1 = q+1$.

El siguiente resultado, inicialmente conjeturado por E. Artin en su tesis y probado por Hasse en la década de 1930, nos dice que este es el valor buscado, con un margen de error de $2\sqrt{q}$. Cabe destacar que la cota y el término de error dependen solo de la elección del cuerpo y no de la curva elíptica.

Teorema 4.5 (Teorema de Hasse)**.** *Sea E una curva elíptica sobre $\mathbb{F}_q$. Entonces*

$$\left| \#E(F_q) - q - 1 \right| \leq 2\sqrt{q}.$$

Para probarlo, precisaremos de algunos resultados previos, que se presentan a continuación. Además, emplearemos el endomorfismo de Frobenius, que recordamos que es la aplicación

$$\begin{aligned} \phi_q : \overline{\mathbb{F}}_q &\longrightarrow \overline{\mathbb{F}}_q \\ x &\longmapsto x^q . \end{aligned}$$

Sea E una curva elíptica sobre $\mathbb{F}_q$. ϕ_q es un endomorfismo de E y actúa sobre los puntos en $E(\overline{\mathbb{F}}_q)$ como $\phi_q(O) = O$ y $\phi_q(x,y) = (x^q, y^q)$, donde estamos escribiendo los puntos de $E(\overline{\mathbb{F}}_q)$ distintos de O como $(x,y) \in \overline{\mathbb{F}}_q^2$. Se tiene, además, la caracterización del siguiente lema:

Lema 4.6. *Sea E una curva elíptica sobre $\mathbb{F}_q$. Denotando por $(x,y) \in \mathbb{F}_q^2$ un punto de $E(\overline{\mathbb{F}}_q)$ distinto del punto del infinito, se tiene que*

(1) $\phi_q(x,y) \in E(\overline{\mathbb{F}}_q)$.

(2) $(x,y) \in E(\mathbb{F}_q)$ *si y solo si* $\phi_q(x,y) = (x,y)$.

Demostración. (1) Dada la ecuación $y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$ con $a_i \in \mathbb{F}_q$, elevando a la potencia q-ésima obtenemos

$$(y^q)^2 + a_1(x^q y^q) + a_3(y^q) = (x^q)^3 + a_2(x^q)^2 + a_4(x^q) + a_6,$$

por lo que $(x^q, y^q) \in E(\overline{\mathbb{F}}_q)$.

(2) El endomorfismo de Frobenius caracteriza $\mathbb{F}_q$, puesto que $x \in \mathbb{F}_q \iff \phi_q(x) = x$. Entonces, dado que $(x,y) \in E(\overline{\mathbb{F}}_q)$, tenemos

$$(x,y) \in E(\mathbb{F}_q) \iff (x,y) \in \mathbb{F}_q \iff \phi_q(x) = x \text{ y } \phi_q(y) = y \iff \phi_q(x,y) = (x,y). \qquad \square$$

La parte (2) de la proposición precedente se puede reformular como sigue: dado un punto $(x,y) \in E(\overline{\mathbb{F}}_q)$ con E una curva elíptica definida sobre $\mathbb{F}_q$, se tiene que $(x,y) \in E(\mathbb{F}_q)$ si y solo si $(x,y) \in \ker(\phi_q - 1)$. Esta caracterización será relevante para la prueba del teorema de Hasse. Los resultados que siguen exploran otras importantes propiedades de la aplicación $\phi_q - 1$.

Proposición 4.7. *Sea E una curva elíptica sobre $\mathbb{F}_q$. Sean r y s los endomorfismos multiplicación por los enteros r y s, no ambos nulos. Entonces, el endomorfismo $r\phi_q + s$ es separable si y solo si $p \nmid s$.*

Demostración. Usando la proposición 3.8, podemos escribir el endomorfismo multiplicación por r en la forma

$$r(x,y) = (f_r(x), y g_r(x))$$

con f_r y g_r funciones racionales con coeficientes en $\overline{\mathbb{F}}_q$. Entonces,

$$(\phi_q r)(x,y) = (f_{r\phi_q}(x), y S_{r\phi_q}(x)) = \left(f_r^q(x), y^q S_r^q(x)\right) = \left(f_r^q(x), y(x^3 + Ax + B)^{(q-1)/2} g_r^q(x)\right).$$

Por consiguiente, $f'_{r\phi_q}/g_{r\phi_q} = q f_r^{q-1} f'_r / g_{r\phi_q} = 0$.

Por otro lado, por la proposición 3.13, se tiene que $f'_s/g_s = s$. Aplicando el lema 3.12,

$$\frac{f'_{r\phi_q+s}}{g_{r\phi_q+s}} = \frac{f'_{r\phi_q}}{g_{r\phi_q}} + \frac{f'_s}{g_s} = 0 + s = s,$$

de donde sigue que $f'_{r\phi_q+s} \neq 0$ si y solo si $p \nmid s$. $\square$

El siguiente corolario es crucial para la demostración del teorema de Hasse. En él usaremos el hecho de que, dado que ϕ_q es un endomorfismo de E, también lo es $\phi_q^m = \phi_q \circ \phi_q \circ \ldots \circ \phi_q$ para todo $m \geq 1$. Adicionalmente, la multiplicación por -1 es un endomorfismo de E, de modo que la suma $\phi_q^m - 1$ también.

Corolario 4.8. *Sea E una curva elíptica sobre $\mathbb{F}_q$ y sea $m \geq 1$.*

(1) $\ker(\phi_q^m - 1) = E(\mathbb{F}_{q^m})$.

(2) $\phi_q^m - 1$ *es un endomorfismo separable de E y* $\#E(\mathbb{F}_{q^m}) = \deg(\phi_q^m - 1)$.

Demostración. (1) sigue del lema 4.6, teniendo en cuenta que el endomorfismo de Frobenius de $\mathbb{F}_{q^m}$ es ϕ_q^m.

(2) es consecuencia de aplicar las proposiciones 4.7 y 3.11. □

Lema 4.9. *Sean r y s como en la proposición 4.7 y sea $a = q+1-\deg(\phi_q-1)$. Entonces,* $\deg(r\phi_q - s) = r^2q + s^2 - rsa$.

Demostración. Puede verse en [10, Lemma 4.8]. □

Tras estos preparativos, estamos ya en condiciones de probar el teorema de Hasse:

Demostración del teorema de Hasse (teorema 4.5). Sea

$$a = q + 1 - \#E(\mathbb{F}_q) = q + 1 - \deg(\phi_q - 1),$$

donde en la última igualdad hemos aplicado el corolario 4.8. Queremos ver que $|a| \leq 2\sqrt{q}$.

Sean r y s como en la proposición 4.7. Puesto que $\deg(r\phi_q - s) \geq 0$, por el lema 4.9 se tiene que

$$q\left(\frac{r}{s}\right)^2 - a\,\frac{r}{s} + 1 \geq 0.$$

Como $\mathbb{Q}$ es denso en $\mathbb{R}$, sigue que $qx^2 - ax + 1 \geq 0$ para todo $x \in \mathbb{R}$. Por tanto, el discriminante de este polinomio es no positivo, es decir, $a^2 - 4q \leq 0$. Concluimos que $|a| \leq 2\sqrt{q}$. □

Capítulo 5

Criptografía de curva elíptica

Desde 1985, cuando el uso de curvas elípticas en criptografía fue propuesto por Neal Koblitz y Victor Miller de forma independiente [23, 27], los sistemas criptográficos basados en curvas elípticas han sido objeto de un amplio estudio. Comenzaron a recibir aceptación comercial a finales de la década de los 90, con la especificación de protocolos por parte de organizaciones de estándares acreditadas y la incorporación de estos protocolos en productos de seguridad por parte de compañías privadas. En la actualidad, se utilizan en protocolos como TLS (Transport Layer Security) [30] y Bitcoin [17], así como en la generación de firmas digitales y otras aplicaciones criptográficas.

En este capítulo, tras una breve introducción al contexto relevante en el ámbito de la criptografía, se describe el problema del logaritmo discreto, que es la base para la criptografía de curva elíptica (ECC, por sus siglas en inglés). Además, se discuten las ventajas de estos y sus principales ataques y vulnerabilidades. A continuación, se analizan el protocolo de Diffie-Hellman en curvas elípticas (ECDH) para el intercambio de claves, el protocolo de criptografía asimétrica de ElGamal y el algoritmo de firma digital con curvas elípticas (ECDSA) para la generación y verificación de firmas digitales. Las fuentes consultadas han sido, fundamentalmente, el libro de Hankeron, Menezes y Vanstone [4] y los libros de Stallings [34] y Silverman [8].

5.1. Introducción a la criptografía de curva elíptica

Consideramos un modelo de comunicación en el que dos entidades, A y B, desean comunicarse de forma segura mediante un canal inseguro, protegiéndose de posibles ataques de un adversario, E.

Definición 5.1. Un sistema de comunicación es *seguro* si garantiza:

- *Confidencialidad*: la información no puede ser leída por entidades no autorizadas.
- *Integridad*: la información solo puede ser modificada de forma autorizada.

- *Autenticación del origen de los datos*: es posible corroborar que los datos recibidos realmente provienen de la entidad que dice haberlos enviado.
- *Autenticación de entidades*: es posible corroborar que las entidades involucradas son quienes dicen ser.
- *No repudio*: una entidad no puede negar haber realizado una acción o aceptado un compromiso.

En ocasiones se requiere además garantizar el *anonimato* de las entidades involucradas y el *control de acceso* a los recursos, es decir, que solo puedan acceder a estos entidades autorizadas.

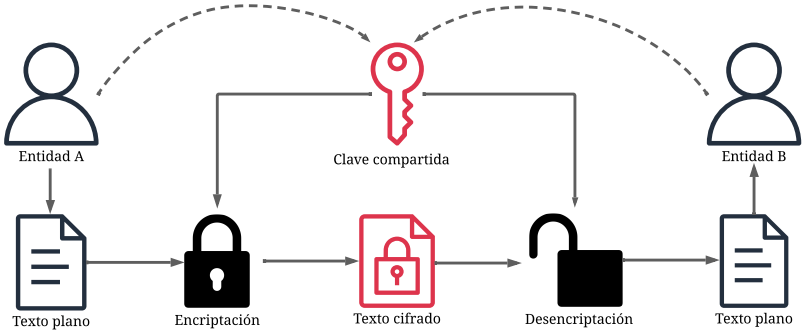


Figura 5.1: Modelo de encriptación simétrica.

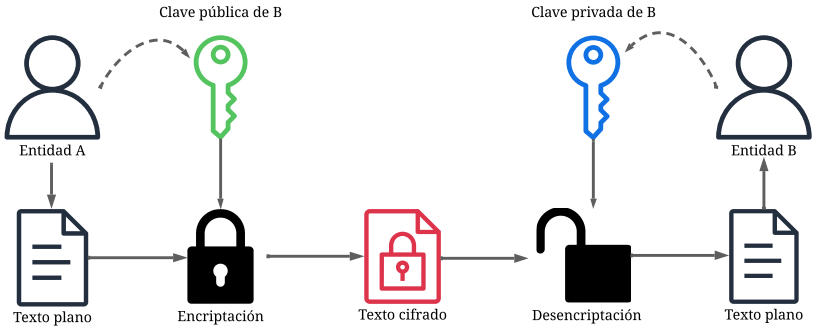


Figura 5.2: Modelo de encriptación asimétrica.

Los sistemas criptográficos se pueden clasificar en dos paradigmas fundamentales: de clave simétrica y de clave asimétrica o pública. En un sistema simétrico, representado en la Figura 5.1, A y B comparten una única clave secreta, independiente de la información que desean comunicar, que utilizan tanto para cifrar como para descifrar los mensajes. En un sistema asimétrico, representado en la Figura 5.2, A y B disponen de sendos pares de claves, cada uno con una clave pública y otra pri-

vada, siendo una de ellas utilizada en el proceso de encriptación y la otra, en el de desencriptación. Además, se deben cumplir las siguientes propiedades:

(P1) Cada clave privada debe ser conocida únicamente por su propietario.

(P2) Debe ser computacionalmente inviable deducir una clave privada a partir de la clave pública.

(P3) Debe ser imposible (o, al menos, computacionalmente inviable) desencriptar un mensaje cifrado con una clave pública determinada si se desconoce la clave privada correspondiente.

(P4) Conocer el algoritmo de cifrado, una clave pública y disponer de mensajes cifrados con ella no puede ser suficiente para deducir la clave privada.

La criptografía asimétrica, que fue introducida por Whitfield Diffie y Martin Hellman en 1976, resuelve los dos principales problemas de la criptografía simétrica: la distribución de claves (i.e., cómo entregar la clave secreta a A y B usando un canal seguro y que garantice autenticidad) y la escalabilidad (en un modelo con N entidades, dado que cada par de entidades requiere una clave secreta propia, el número de claves total es $N(N-1)/2$, que crece de forma cuadrática). No obstante, es importante aclarar una idea errónea bastante extendida: que la criptografía asimétrica es superior a la simétrica. Al ser esta última más costosa computacionalmente, ambos paradigmas se utilizan en conjunto en la práctica, optando por uno u otro según las características y el estado de la comunicación.

La seguridad de los sistemas de criptografía asimétrica depende en gran medida de la propiedad (P2), para la cual es común emplear como base matemática problemas de teoría de números de gran complejidad computacional como la factorización de enteros (la base de RSA, propuesto en 1977, uno de los primeros y más populares esquemas de criptografía asimétrica), el problema del logaritmo discreto (la base de los esquemas de Diffie-Hellman, ElGamal y DSA) o un caso particular de este, el problema del logaritmo discreto en curvas elípticas (la base de los esquemas de criptografía de curva elíptica), que es el que nos ocupa en este trabajo.

No se conocen algoritmos polinómicos (es decir, con tiempo de ejecución del orden $\mathcal{O}((\log_2 N)^k)$ para algún $k > 0$ constante, siendo N la entrada y $\log_2 N$ su número de bits) para ninguno de los tres problemas anteriores, pero sí se han desarrollado algoritmos subexponenciales (con tiempo de ejecución menor a $\mathcal{O}(N^\varepsilon)$ para todo $\varepsilon > 0$) para la factorización de enteros y DLP. No obstante, para el problema del logaritmo discreto en curvas elípticas no se conocen algoritmos subexponenciales, lo cual permite que esquemas criptográficos basados en curvas elípticas alcancen un nivel de seguridad equivalente al de esquemas como RSA con claves de tamaño significativamente menor, como se puede apreciar en el cuadro 5.1. Los consecuentes incrementos en velocidad de computación y disminuciones en el uso de memoria, ancho de banda y energía son algunos de los motivos detrás de la popularidad de la que goza la criptografía de curva elíptica actualmente.

Cuadro 5.1: Comparación entre el número de bits requeridos en la clave de algoritmos simétricos, en las claves públicas (L) y privadas (N) de Diffie-Hellman y DSA, en el módulo de RSA y en el orden n de ECC para obtener un nivel de seguridad equivalente. Se indica también el factor entre los tamaños de RSA y ECC.

Algoritmos de clave simétrica	**Diffie-Hellman, DSA**	**RSA**	**ECC**	**Factor RSA/ECC**
80	L = 1024, N = 160	1024	160	6.4
112	L = 2048, N = 224	2048	224	$64/7 \approx 9.143$
128	L = 3072, N = 256	3072	256	12
196	L = 7680, N = 384	7680	384	20
256	L = 15360, N = 512	15360	512	30

5.2. El problema del logaritmo discreto en curvas elípticas

Definición 5.2. Sea $G = \langle g \rangle$ un grupo cíclico finito de orden n. El *problema del logaritmo discreto* (DLP, *discrete logarithm problem*) en G consiste en, dado $y \in G$, determinar $x \in \{0, 1, \dots, n-1\}$ tal que $y = g^x$.

Con la notación de la definición anterior, un sistema criptográfico basado en el problema del logaritmo discreto sobre un grupo G debe ser tal que encontrar x sea computacionalmente inviable, pero que calcular la operación del grupo sea sencillo. La elección de G es importante: por ejemplo, en el grupo aditivo $\mathbb{F}_p^+$ se puede utilizar el algoritmo de Euclides para hallar el valor de m en la ecuación lineal $b = ma \bmod p$, para $a, b \in \mathbb{F}_p^+$ dados. Dicho algoritmo requiere a lo sumo $2 \log n$ pasos, por lo que $\mathbb{F}_p^+$ no da lugar a un problema suficientemente robusto.

El caso en el que G es un subgrupo del grupo multiplicativo de un cuerpo finito, $G < \mathbb{F}_p^\times$ (que necesariamente implica que G es cíclico [11, Proposition 3.6.14]), es la versión utilizada en las versiones originales de los algoritmos de Diffie-Hellman, ElGamal y DSA.

En el caso de que G sea un subgrupo cíclico del grupo de puntos de una curva elíptica, hablamos del *problema del logaritmo discreto en curvas elípticas*. Existen adaptaciones de los algoritmos de Diffie-Hellman, ElGamal y DSA al ECDLP, que veremos en las secciones 5.4, 5.5 y 5.6.

Definición 5.3. Consideremos una curva elíptica E definida sobre un cuerpo finito $\mathbb{F}_q$, un punto $P \in E$ de orden n, y un punto $Q \in \langle P \rangle$. El *problema del logaritmo discreto en curvas elípticas* (ECDLP, *elliptic curve discrete logarithm problem*) consiste en determinar $l \in \{0, 1, \dots, n-1\}$ tal que $Q = lP$. l se denomina *logaritmo discreto de Q en base P* y se denota $l = \log_P Q$.

En la criptografía de curva elíptica (ECC, *elliptic curve cryptography*), Q es la clave pública y l, la clave privada.

Un ataque de fuerza bruta al ECDLP calcularía la secuencia P, $2P$, $3P$... hasta encontrar Q. En el peor caso, necesitaría n iteraciones o pasos, con una media de $n/2$ iteraciones. El ataque más eficiente conocido es una combinación del *algoritmo de Pohlig-Hellman* y del *algoritmo ρ de Pollard*, y tiene complejidad $\mathcal{O}(\sqrt{p})$, donde p es el mayor divisor primo de n. Generalmente, n es primo y aproximadamente igual al orden del cuerpo finito $\mathbb{F}_q$ sobre el que está definido la curva elíptica, por lo que la complejidad es $\mathcal{O}(\sqrt{q})$, que es exponencial en $\log_2 q$. Esta es muy superior a la de los mejores algoritmos disponibles para el DLP sobre grupos $G < \mathbb{F}_p^*$, conocidos como *métodos de index calculus*, con un tiempo subexponencial de orden $\mathcal{O}(\exp\left(c\sqrt[3]{(\log q)(\log\log q)^2}\right)$.

El tiempo de ejecución de los algoritmos que emplean curvas elípticas está principalmente determinado por la operación de "multiplicación escalar" lP. Para un cálculo eficiente de la misma, se puede utilizar la versión aditiva del algoritmo de exponenciación binaria, que se conoce como *double and add*, u otras variantes del mismo. Este se basa en emplear la representación binaria de l y recorrerla de izquierda a derecha. Partiendo de O (el punto del infinito), por cada 0 se dobla y por cada 1 se dobla y se suma de nuevo el punto original. De esta forma, se reduce el total de l operaciones de grupo a $\log_2 l$ iteraciones de *double and add*, cuyos detalles se exponen en el algoritmo 1.

Algoritmo 1 Algoritmo de *double and add* para la multiplicación escalar.

Entrada: $l = l_{t-1}2^{t-1} + \cdots + l_1 2 + l_0$, $l_i \in \{0,1\}$, $i = 0, \ldots, t-1$, E curva elíptica sobre $\mathbb{F}_q$, $P \in E$.
Salida: lP.

1: $Q \leftarrow O$.
2: **for** i de $t-1$ a 0 en orden decreciente **do**
3: $\quad Q \leftarrow 2Q$
4: $\quad$ **if** $l_i = 1$ **then**
5: $\quad\quad Q \leftarrow Q + P$.
6: $\quad$ **end if**
7: **end for**
8: **Devolver:** Q.

5.3. Elección de los parámetros de dominio

Aunque los algoritmos más rápidos para resolver el ECDLP en general para una curva elíptica E sobre $\mathbb{F}_q$ son exponenciales, existen casos particulares para los que ciertos ataques permiten definir algoritmos subexponenciales que resuelven el ECDLP. Con el objetivo de asegurar la resistencia a todos los ataques conocidos, es necesario seleccionar cuidadosamente los parámetros de dominio acordados entre las entidades participantes.

Definición 5.4. En la criptografía de curva elíptica, los *parámetros de dominio* $D = (q, a, b, P, n, h)$ son:

(1) q, el orden del cuerpo finito $\mathbb{F}_q$. Generalmente se elige un orden primo o 2^m para un entero $m \geq 2$.

(2) Dos coeficientes $a, b \in \mathbb{F}_q$ que determinan la curva elíptica E sobre $\mathbb{F}_q$ a emplear: $y^2 = x^3 + ax + b$, si q es un primo distinto de 2 y 3, o $y^2 + xy = x^3 + ax^2 + b$ si char $\mathbb{F}_q = 2$.

(3) Un *punto base* $P \in E(\mathbb{F}_q)$ distinto del punto del infinito.

(4) El orden n de P, que debe ser primo.

(5) El *cofactor* $h = \#E(\mathbb{F}_q)/n$.

Para evitar los ataque de Pohlig-Hellman y el ataque ρ de Pollard (cuyos detalles, en parte de naturaleza probabilística, se incluyen en el anexo II), $\#E(\mathbb{F}_q)$ debe ser divisible por un orden primo n suficientemente grande. Como mínimo, se requiere que $n > 2^{160}$. La seguridad es todavía mayor cuando se elige E de forma que $\#E(\mathbb{F}_q)$ sea primo o *casi primo*, esto es, que el cofactor h sea pequeño (por ejemplo, $h = 1, 2, 3$ o 4). Nótese que si $h > 1$, algunos algoritmos requieren de pasos adicionales para asegurar su correcto funcionamiento [28].

Otros ataques a resistir son los *ataques basados en isomorfismos*. Mencionamos algunos aspectos importantes: para evitar ataques en curvas *anómalas*, E debe cumplir que $\#E(\mathbb{F}_q) \neq q$. Para hacer frente a los *ataques del emparejamiento de Weil y de Tate*, n no debe dividir $q^k - q$ para ningún $k \in [1, C]$ entero, donde C es suficientemente grande (si $n > 2^{160}$, basta comprobarlo hasta $C = 20$). Con el objetivo de resistir el *ataque del descenso de Weil*, no se deben emplear cuerpos finitos de orden 2^m si m no es primo.

Para cumplir con varias de estas condiciones, es necesario saber con exactitud el número de puntos de la curva elíptica, $\#E(\mathbb{F}_q)$. El teorema de Hasse 4.5 nos proporciona solamente una cota, pero el *algoritmo de Schoof*, que se basa en la aproximación dada por Hasse, permite calcular el valor exacto de $\#E(\mathbb{F}_q)$ en tiempo polinómico: emplea $\mathcal{O}((\log q)^8)$ pasos. Sus detalles se pueden consultar en [8, Section XI.3].

Las curvas elípticas supersingulares también deben evitarse, pues se ha probado que son más débiles para el ECDLP que el caso general [15].

No obstante, la generación de parámetros de dominio tiene una implementación delicada y es computacionalmente costosa, por lo que es común emplear parámetros de dominio publicados por autoridades reconocidas como el Instituto Nacional de Normas y Tecnología (NIST, *National Institute of Standards and Technology*) [29], el Grupo de Estándares para la Criptografía Eficiente (SECG, *Standards for Efficient Cryptography Group*) [35] o *ECC Brainpool* [26].

5.4. Elliptic-curve Diffie-Hellman (ECDH)

Como primer ejemplo de aplicación de la ECC, veamos la adaptación del clásico algoritmo de intercambio de claves de Diffie-Hellman al caso de curvas elípticas.

Supongamos que dos entidades A y B desean establecer una clave secreta compartida. Tras fijar los parámetros de dominio, A elige un número entero $n_A \in [1, n-1]$, que será su clave privada, y calcula $P_A = n_A P \in E$, su clave pública. B hace lo análogo, con un entero $n_B \in [1, n-1]$ y $P_B = n_B P \in E$ sus claves privada y pública, respectivamente.

Finalmente, A envía P_A a B, y B envía P_B a A. Ambos calculan la clave secreta compartida $K = n_A P_B = n_B P_A = n_A n_B P \in E$.

La seguridad de este proceso, que se recoge en el algoritmo 2, subyace en la dificultad para un atacante de obtener k, para lo cual tendría que calcular n_A o n_B a partir de P_A o P_B, que es un caso particular del ECDLP.

Algoritmo 2 Intercambio de claves con Elliptic Curve Diffie-Hellman (ECDH)

Entrada: Un punto base P de orden n en una curva elíptica E sobre un cuerpo finito $\mathbb{F}_q$.
Salida: Clave secreta compartida K.

1: A elige su clave privada $n_A \in \{1, 2, \dots, n-1\}$ y B elige su clave privada $n_B \in \{1, 2, \dots, n-1\}$.
2: A calcula su clave pública $P_A = n_A P$ y B calcula su clave pública $P_B = n_B P$.
3: A envía P_A a B y B envía P_B a A.
4: A calcula $K = n_A P_B = n_A n_B P$ y B calcula $K = n_B P_A = n_B n_A P$.
5: **Resultado:** A y B comparten la clave secreta $K = n_A n_B P$.

La clave secreta K se puede utilizar para generar una clave de cifrado simétrico para algoritmos como AES. En tal caso, es común tomar una de las coordenadas de K o una función sencilla de K.

Nótese que ECDH es vulnerable frente a ataques de intermediario, por el mismo motivo que en el algoritmo clásico de Diffie-Hellman. No obstante, la autenticación de las claves de ambas partes evita este problema.

5.5. Sistema de criptografía asimétrica de ElGamal con curvas elípticas

Aunque Diffie-Hellman es un protocolo basado en la criptografía asimétrica, no permite transmitir mensajes específicos escogidos por alguna de las entidades participantes, pues el valor secreto que comparte entre estas no está determinado a priori. Lo mismo ocurre para su versión en curvas

elípticas.

ElGamal propone en 1985 un sistema criptográfico asimétrico basado en el DLP sobre $\mathbb{F}_q^\times$. En esta sección presentamos su variante en curvas elípticas.

Algoritmo 3 Encriptación de ElGamal sobre curvas elípticas

Entrada: Un punto base P en una curva elíptica E sobre un cuerpo finito $\mathbb{F}_q$ y un mensaje m.
Salida: Mensaje cifrado (C_1, C_2).

1: Se elige una clave privada $d \in \{1, 2, \dots, n-1\}$ y se calcula la clave pública $Q = dP$.
2: Se representa m como un punto $M \in E(\mathbb{F}_q)$.
3: Se elige un entero $k \in \{1, 2, \dots n-1\}$ aleatoriamente.
4: Se calcula $C_1 = kP = (x_1 : y_1 : 1)$.
5: Se calcula $C_2 = M + kQ$.
6: **Resultado:** (C_1, C_2).

Algoritmo 4 Desencriptación de ElGamal sobre curvas elípticas

Entrada: Un punto base P en una curva elíptica E sobre un cuerpo finito $\mathbb{F}_q$, un mensaje cifrado (C_1, C_2) y la clave privada d correspondiente a la clave pública con la que se cifró.
Salida: Mensaje en texto plano m.

1: Se calcula $M = C_2 - dC_1$ y se extrae m de M.
2: **Resultado:** m.

Con la notación de los algoritmos 3 y 4, obsérvese que el desafío computacional es averiguar kQ sabiendo únicamente Q y C_1, que es equivalente al problema planteado en ECDH.

No hay una forma natural de asignar un punto $M \in E(\mathbb{F}_q)$ a un mensaje m. No obstante, existe una variante de este procotolo, MV-ElGamal (por sus creadores, Menezes y Vanstone), que especifica un método para llevar a cabo esta asignación. Sus detalles se pueden consultar en [8, Chapter XI].

5.6. Elliptic Curve Digital Signature Algorithm (ECDSA)

El objetivo de los *esquemas de firmas digitales* es la autenticación del origen de los datos, la integridad de los datos y/o el no repudio. Estos esquemas, que forman parte de los sistemas de criptografía asimétrica, permiten que una entidad A genere una firma digital s para un determinado mensaje m empleando su clave privada d. Otra entidad B que reciba m y s puede verificar que s fue generado por A para el mensaje m empleando la clave pública de A, Q. La Figura 5.3 muestra una representación de este modelo.

Puesto que el proceso de verificación solo necesita emplear s, m y Q, puede ser realizado por una entidad externa para impedir que A niegue haber firmado m. Como s depende del mensaje m, la

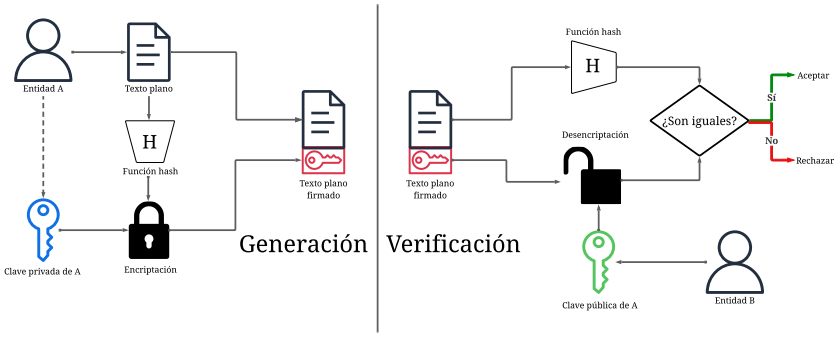


Figura 5.3: Modelo para la generación y verificación de firmas digitales.

firma sería invalidada para otro mensaje m', con lo que un atacante que pretendiera hacer ver que A también firmó m' fallaría.

El Algoritmo de Firma Digital de Curva Elíptica (ECDSA, *Elliptic Curve Digital Signature Algorithm*) adapta el clásico Algoritmo de Firma Digital (DSA, *Digital Signature Algorithm*) al grupo de curvas elípticas. ECDSA es el algoritmo de firma con curvas elípticas más estandarizado [4], pero existen otros, también empleados en diversos estándares, como EC-KCDSA o EdDSA [21].

Para presentar el protocolo, distinguimos entre el proceso de generación y de verificación de firmas. En ambos algoritmos se empleará el concepto de *función hash*, cuyo propósito es transformar datos de entrada en una salida corta que sirva a modo de identificador de esos datos, y que tenga una baja probabilidad de ser la salida también para datos de entrada distintos.

Definición 5.5. Una *función hash* H es una aplicación que lleva una cadena de texto de longitud arbitraria en una cadena de texto de longitud fija. Su nivel de seguridad depende de su cumplimiento de las siguientes tres propiedades:

1. *Resistencia a la preimagen*: dada una imagen h de H, es computacionalmente inviable encontrar una preimagen x tal que $H(x) = h$.
2. *Segunda resistencia a la preimagen*: dada una cadena de entrada x, es computacionalmente inviable encontrar otra cadena de entrada x', con $x \neq x'$, tal que $H(x) = H(x')$.
3. *Resistencia a la colisión*: es computacionalmente inviable encontrar cadenas de entrada x y x', con $x \neq x'$, tales que $H(x) = H(x')$.

La propiedad de resistencia a la colisión implica la segunda resistencia a la preimagen, pero no la resistencia a la preimagen. Una función hash que no satisface alguna de estas tres propiedades es vulnerable a ataques maliciosos.

En lo que sigue, H denotará una función hash con a lo sumo n bits de salida (si la salida fuese mayor, basta truncarla a n bits). Cabe destacar que si H no cumple las tres propiedades de seguridad de las funciones hash, ECDSA es vulnerable a que un atacante que consiga firmas de mensajes

generadas por una entidad A pueda forjar firmas válidas para nuevos mensajes en nombre de A [8], por lo que es recomendable usar algoritmos estándar como SHA-2 o SHA-3.

Algoritmo 5 Generación de firmas con ECDSA

Entrada: Un punto base P de orden n en una curva elíptica E sobre $\mathbb{F}_q$ y un mensaje m.

Salida: Una firma (s_1, s_2) para el mensaje m.

1: Se elige una clave privada $d \in \{1, 2, \dots, n-1\}$ y se calcula la clave pública $Q = dP$.
2: Se elige un entero $k \in \{1, 2, \dots n-1\}$ aleatoriamente.
3: Se calcula $kP = (x_1 : y_1 : 1)$ y se toma un representante entero $\overline{x_1}$ de $x_1 \in \mathbb{F}_q$.
4: Se calcula $s_1 = \overline{x_1} \pmod n$. Si $s_1 = 0$, se vuelve al paso 2.
5: Se calcula $e = H(m)$.
6: Se calcula $s_2 = k^{-1}(e + dr) \pmod n$. Si $s_2 = 0$, se vuelve al paso 3.
7: **Resultado:** La firma es el par (s_1, s_2).

Algoritmo 6 Verificación de firmas con ECDSA

Entrada: Un punto base P de orden n en una curva elíptica E sobre $\mathbb{F}_q$, un mensaje m, una firma (s_1, s_2) de m la clave pública Q correspondiente a la clave privada que se usó para generar (s_1, s_2).

Salida: La aceptación o el rechazo de la firma.

1: Si $s_1 \notin [1, n-1] \cap \mathbb{Z}$ o $s_2 \notin [1, n-1] \cap \mathbb{Z}$, la firma se rechaza.
2: Se calcula $e = H(m)$.
3: Se calcula $w = s_2^{-1} \pmod n$.
4: Se calcula $u_1 = ew \pmod n$ y $u_2 = s_1 w \pmod n$.
5: Se calcula $X = u_1 P + u_2 Q$. Si X es el punto del infinito, la firma se rechaza.
6: Se toma un representante entero $\overline{x_1}$ de x_1, donde $X = (x_1 : y_1 : 1)$, y se calcula $v = \overline{x_1} \pmod n$.
7: Si $v = s_1$, la firma se acepta. En otro caso, se rechaza.

Demostración de que la verificación de firmas con ECDSA es correcta. Queremos ver que si una entidad A genera una firma (s_1, s_2) con el algoritmo 5, entonces el algoritmo 6, ejecutado por otra entidad B, devuelve el resultado correcto.

Por definición de las variables empleadas en 6, el punto X calculado por B es

$$X = u_1 P + u_2 Q \equiv ewP + s_1 wdP \equiv s_2^{-1}(e + s_1 d)P \pmod n.$$

Ahora bien, si la firma fue generada correctamente, se cumple que $s_2 \equiv k^{-1}(e + ds_1) \pmod n$. Equivalentemente, $k \equiv s_2^{-1}(e + s_1 d) \pmod n$. Por consiguiente, $X = kP$ y se tiene que $v = s_1$. □

Las comprobaciones de que s_1 y s_2 estén en el intervalo $[1, n-1]$ en la verificación permiten evitar ataques que focalizan este punto.

Anexo I

El invariante j a través de transformaciones proyectivas

Este anexo está dedicado a probar que el invariante j de una curva elíptica no cambia bajo transformaciones proyectivas que llevan dicha curva a otra curva elíptica. Esta demostración no se incluye en el texto principal por requerir la introducción de conceptos adicionales de geometría proyectiva (las transformaciones proyectivas y la razón doble) que no se utilizan en el resto del trabajo, y por precisar cálculos relativamente pesados y no demasiado ilustrativos.

Como en la sección 2.3, a lo largo de este anexo supondremos que $\operatorname{char} K \neq 2$ y seguiremos, principalmente, el libro de Garrity et al. [2].

En primer lugar, definimos los *cambios de coordenadas lineales proyectivos*, para lo cual necesitamos introducir el *grupo lineal proyectivo*. Para una profundización al respecto, recomendamos consultar [25].

Definición I.1. El *grupo lineal proyectivo* $\mathrm{PGL}_n(K)$ es el grupo cociente

$$\mathrm{PGL}_n(K) = \frac{\mathrm{GL}_n(K)}{Z(K)},$$

donde $\mathrm{GL}_n(K)$ es el grupo lineal general, formado por las matrices $n \times n$ invertibles con entradas en K, y $Z(K)$ es su centro, esto es, el subgrupo normal de matrices escalares (que es isomorfo a $K^\times$).

Los elementos de $\mathrm{PGL}_n(K)$ reciben el nombre de *transformaciones proyectivas* o *cambios de coordenadas lineales proyectivos*.

Observación I.2. Sigue de la definición anterior que las transformaciones proyectivas son matrices consideradas iguales cuando difieren en un factor $\lambda \in K^\times$.

Nótese que la acción canónica de $\mathrm{GL}_{n+1}(K)$ sobre K^{n+1} induce una acción de $\mathrm{GL}_{n+1}(K)$ sobre

$\mathbb{P}^n$. Es inmediato ver que las matrices escalares actúan como la identidad y que son el núcleo de la acción, por lo que es razonable cocientar por ellas para obtener una acción de $\mathrm{PGL}_{n+1}(K)$ sobre $\mathbb{P}^n$.

Por otro lado, una curva elíptica en forma de Legendre está dada en la forma $y^2 = f(x)$, con $f \in K[x]$ un polinomio cúbico. Homogeneizando, obtenemos que su clausura proyectiva está dada por la ecuación $zy^2 = \widehat{f}(x,z)$. Consideremos entonces la recta proyectiva $\mathbb{P}^1$ con coordenadas $(x : z)$.

Nuestro primer objetivo es probar que, si una transformación proyectiva permuta el punto del infinito $(1 : 0)$ y dos de las raíces de f, $(0 : 1)$ y $(1 : 1)$, podemos decir cuáles son las posibles imágenes de la tercera raíz, $(\lambda : 1)$. Para ello, utilizaremos el concepto de *razón doble*.

Definición I.3. La *razón doble* de cuatro puntos distintos $p_1 = (x_1 : z_1)$, $p_2 = (x_2 : z_2)$, $p_3 = (x_3 : z_3)$, $p_4 = (x_4 : z_4) \in \mathbb{P}^1$ está dada por

$$[p_1, p_2, p_3, p_4] = \frac{(x_2z_4 - z_2x_4)(x_1z_3 - x_3z_1)}{(x_1z_4 - z_1x_4)(x_2z_3 - x_3z_2)}. \tag{I.1}$$

Ejemplo I.4. La razón doble de $p_1 = (1 : 0)$, $p_2 = (0 : 1)$, $p_3 = (1 : 1)$ y $p_4 = (\lambda : 1)$ es $[p_1, p_2, p_3, p_4] = \lambda$.

El origen de la razón doble está en el hecho de que dados tres puntos p_1, p_2, $p_3 \in \mathbb{P}^1$ distintos entre sí y otros tres puntos q_1, q_2, $q_3 \in \mathbb{P}^1$ distintos entre sí, siempre existe una transformación proyectiva $T \in \mathrm{PGL}_2(K)$ tal que $T(p_1) = q_1, T(p_2) = q_2$ y $T(p_3) = q_3$ [2]. No obstante, si se considera también un cuarto punto $p_4 \in \mathbb{P}^1$, su imagen por T está determinada por la elección de las imágenes de p_1, p_2 y p_3. Es decir, una colección de cuatro puntos en $\mathbb{P}^1$ tiene una geometría intrínseca que no depende de la elección de coordenadas, y esta geometría se puede capturar a través de la razón doble. Esta propiedad se expresa formalmente en la siguiente proposición.

Proposición I.5. *Los transformaciones proyectivas de $\mathbb{P}^1$ preservan la razón doble.*

Demostración. Sea $T \in \mathrm{PGL}_2(K)$. T se corresponde con una aplicación de la forma $T(x,z) = (ax + bz : cx + dz)$ con $ad - bc \neq 0$. Veamos que $[T(p_1), T(p_2), T(p_3), T(p_4)] = [p_1, p_2, p_3, p_4]$.

Por definición, $[T(p_1), T(p_2), T(p_3), T(p_4)]$ es igual a

$$\frac{((ax_2+bz_2)(cx_4+dz_4)-(cx_2+dz_2)(ax_4+bz_4))((ax_1+bz_1)(cx_3+dz_3)-(cx_1+dz_1)(ax_3+bz_3))}{((ax_1+bz_1)(cx_4+dz_4)-(cx_1+dz_1)(ax_4+bz_4))((ax_2+bz_2)(cx_3+dz_3)-(cx_2+dz_2)(ax_3+bz_3))}.$$

Ahora bien, $((ax_2 + bz_2)(cx_4 + dz_4) - (cx_2 + dz_2)(ax_4 + bz_4))$ es igual a $(ac - ac)x_2x_4 + (ad - bc)x_2z_4 + (bc - ad)x_4z_2 + (bd - bd)z_2z_4 = (ad - bc)(x_2z_4 - x_4z_2)$.

Análogamente,

$$((ax_1 + bz_1)(cx_3 + dz_3) - (cx_1 + dz_1)(ax_3 + bz_3)) = (ad - bc)(x_1z_3 - x_3z_1)$$
$$((ax_1 + bz_1)(cx_4 + dz_4) - (cx_1 + dz_1)(ax_4 + bz_4)) = (ad - bc)(x_1z_4 - x_4z_1)$$
$$((ax_2 + bz_2)(cx_3 + dz_3) - (cx_2 + dz_2)(ax_3 + bz_3)) = (ad - bc)(x_2z_3 - x_3z_2).$$

Por tanto,

$$[T(p_1), T(p_2), T(p_3), T(p_4)] = \frac{(x_2z_4 - x_4z_2)(x_1z_3 - x_3z_1)}{(x_1z_4 - x_4z_1)(x_2z_3 - x_3z_2)} = [p_1, p_2, p_3, p_4].$$

□

Aunque la razón doble depende del orden de los puntos considerados y hay $4! = 24$ permutaciones de cuatro puntos, no todas ellas dan valores diferentes. Tomando $\lambda := [p_1, p_2, p_3, p_4]$ y calculando explícitamente las razones dobles para las diferentes permutaciones, se ve que

$$\begin{aligned}
\lambda &= [p_1, p_2, p_3, p_4] = [p_2, p_1, p_4, p_3] = [p_3, p_4, p_1, p_2] = [p_4, p_3, p_2, p_1] \\
1 - \lambda &= [p_1, p_3, p_2, p_4] = [p_2, p_4, p_1, p_3] = [p_3, p_1, p_4, p_2] = [p_4, p_2, p_3, p_1] \\
\frac{\lambda}{\lambda - 1} &= [p_1, p_4, p_3, p_2] = [p_2, p_3, p_4, p_1] = [p_3, p_2, p_1, p_4] = [p_4, p_1, p_2, p_3] \\
\frac{1}{\lambda} &= [p_1, p_2, p_4, p_3] = [p_2, p_1, p_3, p_4] = [p_3, p_4, p_2, p_1] = [p_4, p_3, p_1, p_2] \\
\frac{1}{1 - \lambda} &= [p_1, p_3, p_4, p_2] = [p_2, p_4, p_3, p_1] = [p_3, p_1, p_2, p_4] = [p_4, p_2, p_1, p_3] \\
\frac{\lambda - 1}{\lambda} &= [p_1, p_4, p_2, p_3] = [p_2, p_3, p_1, p_4] = [p_3, p_2, p_4, p_1] = [p_4, p_1, p_3, p_2].
\end{aligned}$$

Por tanto, como también se deduce de la simetría de la expresión que define la razón doble, I.1, la razón doble es invariante al:

- permutar simultáneamente la primera y segunda pareja de los cuatro puntos,
- permutar estas parejas entre sí, y
- combinar las dos anteriores permutaciones.

Sean $\lambda, \overline{\lambda} \in \overline{K}$ tales que $\lambda, \overline{\lambda} \neq 0, 1$. Consideremos una transformación proyectiva T que lleve una curva elíptica en forma de Legendre E_λ en otra curva elíptica en forma de Legendre $E_{\overline{\lambda}}$. Se puede demostrar que esto implica que T debe llevar los puntos $(1 : 0)$, $(0 : 1)$, $(1 : 1)$ y $(\lambda : 1)$ en alguna permutación de $(1 : 0)$, $(0 : 1)$, $(1 : 1)$ y $(\overline{\lambda} : 1)$ [6, Chapter 10]. El razonamiento anterior, la proposición I.5 y el ejemplo I.4 nos permiten deducir que

$$\overline{\lambda} \in \left\{ \lambda, 1 - \lambda, \frac{\lambda}{\lambda - 1}, \frac{1}{\lambda}, \frac{1}{1 - \lambda}, \frac{\lambda - 1}{\lambda} \right\}.$$

Así, concluimos que el invariante j es un invariante de la clase de curvas que pueden ser transformadas en E_λ a través de una transformación proyectiva, y que por tanto clasifica las curvas elípticas.

Anexo II

Ataques al problema del logaritmo discreto

Este apéndice está dedicado a exponer el ataque más efectivo al problema del logaritmo discreto en curvas elípticas (ECDLP, por sus siglas en inglés) que se conoce. Este es una combinación de dos algoritmos, el algoritmo de Pohlig-Hellman y el algoritmo ρ de Pollard. Por tener estos un carácter relativamente técnico y propio de la criptografía, se han incluido en un anexo con el fin de no interrumpir la exposición del cuerpo principal del trabajo. Para seguirlo, sugerimos [8, Section XI.5] y [4, Section 4.1].

Se empleará notación introducida en las secciones 5.1, 5.2 y 5.3, por lo que se recomienda haber revisado estos apartados previamente.

A lo largo de este anexo, consideraremos una curva elíptica E definida sobre un cuerpo finito $\mathbb{F}_q$, un punto $P \in E$ de orden n y un punto $Q \in E$ tal que $Q = kP$ para un cierto $k \in \{0, 1, \dots, n-1\}$. Nuestro objetivo será resolver el ECDLP, es decir, encontrar $k = \log_P Q$ dados E, P y Q.

II.1. Algoritmo de Pohlig-Hellman

A diferencia del algoritmo ρ de Pollard, el algoritmo de Pohlig-Hellman no busca calcular $k = \log_P Q$ directamente, sino que su objetivo es reducir el cálculo de k al cálculo de logaritmos discretos en los subgrupos de $\langle P \rangle$ de orden primo.

Sea $n = p_1^{e_1} p_2^{e_2} \cdots p_m^{e_m}$ la factorización de n en primos p_i distintos dos a dos y enteros positivos e_i. El algoritmo de Pohlig-Hellman consiste en calcular $k_i \equiv k \pmod{p_i^{e_i}}$ para cada $i = 1, \dots, m$ y luego

emplear el teorema chino de los restos para resolver el sistema de congruencias

$$k \equiv k_i \pmod{p_i^{e_i}} \quad \text{para } i = 1, \dots, m.$$

y obtener k.

La clave del algoritmo está en que cada k_i se puede obtener calculando e_i logaritmos discretos en el subgrupo de orden p_i de (P). Si todos los p_i son primos pequeños, entonces los correspondientes logaritmos discretos se pueden obtener rápidamente en comparación con el ECDLP original.

Fijemos un k_i y consideremos su representación en base p_i:

$$k_i = z_{i,0} + z_{i,1}p_i + z_{i,2}p_i^2 + \cdots + z_{i,e_i-1}p_i^{e_i-1}, \tag{II.1}$$

con $z_{i,j} \in \{0, 1, \dots, p_i - 1\}$ para $j = 0, \dots, e_i - 1$.

Los valores de $z_{i,0}, z_{i,1}, \dots, z_{i,e_i-1}$, que determinan k_i, se obtienen secuencialmente. En primer lugar, se calcula $P_{i,0} = (n/p_i)P$ y $Q_{i,0} = (n/p_i)Q$. Se tiene entonces que

$$Q_{i,0} = \frac{n}{p_i}Q = k\left(\frac{n}{p_i}P\right) = kP_{i,0} = z_{i,0}P_{i,0},$$

donde en la última igualdad hemos usado que $P_{i,0}$ es un punto de orden p_i, dado que $p_iP_{i,0} = nP$. Por tanto, $z_{i,0} = \log_{P_{i,0}} Q_{i,0}$ y $z_{i,0}$ se puede obtener resolviendo el ECDLP en el subgrupo $(P_{i,0})$, que tiene orden p_i.

A continuación, se calcula $Q_{i,1} = (n/p_i^2)(Q - z_{i,0}P)$. Entonces,

$$\begin{aligned} Q_{i,1} &= \frac{n}{p_i^2}(Q - z_{i,0}P) = \frac{n}{p_i^2}(k - z_{i,0})P = (k - z_{i,0})\left(\frac{n}{p_i^2}P\right) \\ &= (z_{i,0} + z_{i,1}p_i - z_{i,0})\left(\frac{n}{p_i^2}P\right) = z_{i,1}\left(\frac{n}{p_i}P\right) = z_{i,1}P_{i,0}, \end{aligned}$$

donde hemos usado que el orden de $(n/p_i^2)P$ es p_i^2. Obtenemos así que $z_{i,1} = \log_{P_{i,0}} Q_{i,1}$, y que $z_{i,1}$ se puede calcular resolviendo una instancia del ECDLP en el subgrupo $(P_{i,0})$.

Conociendo el valor de $z_{i,0}, z_{i,1}, \dots, z_{i,t-1}$, se puede calcular $z_{i,t} = \log_{P_{i,0}} Q_{i,t}$, con $t \in \{0, \dots, e_i - 1\}$, de forma análoga, donde

$$Q_{i,t} = \frac{n}{p_i^{t+1}}(Q - z_{i,0}P - z_{i,1}p_iP - \cdots - z_{i,t-1}p_i^{t-1}P).$$

Por tanto, para determinar k_i basta resolver e_i instancias del ECDLP en el subgrupo $(P_{i,0})$ y realizar las e_i multiplicaciones escalares de la expresión (II.1). Repitiendo este proceso para cada $i = 1, \dots, m$ y aplicando el teorema chino de los restos, se llega al valor de k.

Por consiguiente, la seguridad de la elección de los parámetros de dominio depende del mayor primo que divide a n. Para resistir este ataque, conviene elegir dichos parámetros de forma que n sea divisible por un primo grande.

II.2. Algoritmo ρ de Pollard

El algoritmo ρ de Pollard es un algoritmo probabilístico que permite resolver el ECDLP en tiempo $O(\sqrt{n})$. Combinándolo con el algoritmo de Pohlig-Hellman, se obtiene un ataque al ECDLP de complejidad $O(\sqrt{p})$, siendo p el mayor primo que divide a n.

El algoritmo ρ de Pollard se basa en la idea de buscar colisiones en una secuencia de puntos generada por P y Q. Específicamente, se busca encontrar dos parejas de enteros (c', d') y (c'', d'') distintas tales que

$$c'P + d'Q = c''P + d''Q.$$

Conociendo (c', d') y (c'', d''), se deduce que

$$(c' - c'')P = (d'' - d')Q = (d'' - d')kP,$$

lo que implica que

$$c' - c'' = (d'' - d')k \pmod{n}.$$

Por consiguiente,

$$k = \log_P Q = (c' - c'')(d'' - d')^{-1} \pmod{n},$$

suponiendo que $d'' - d'$ es invertible en $\mathbb{Z}/n\mathbb{Z}$. En aplicaciones prácticas, n suele ser primo (entre otros motivos, porque ayuda a resistir el ataque de Pohlig-Hellman), por lo que esta no es una restricción importante.

El método de Pollard, así como sus variantes, se fundamenta en el siguiente teorema:

Teorema II.1. *Sea S un conjunto finito de N elementos y sea $f : S \longrightarrow S$ una aplicación. Consideramos la secuencia de elementos de S definida por*

$$x_0 \in S, \quad x_i = f(x_{i-1}) = \underbrace{f \circ f \circ \cdots \circ f}_{i \text{ veces}}(x_0) \quad \text{para } i = 1, 2, \ldots$$

Denotamos por T el mayor entero tal que x_{T-1} aparece una sola vez en la secuencia $(x_i)_{i\geq 0}$, y por L el menor entero tal que $X_{T+L} = x_T$. Es decir, L es el período con el que se repite la secuencia a partir de x_T (véase la Figura II.1). Es común referirse a T como la longitud de la cola y a L como la longitud del ciclo de la secuencia.

Entonces, se cumple que

1. *Existe un índice $1 \leq i < T + L$ tal que $x_{2i} = x_i$.*
2. *Si $f : S \longrightarrow S$ es una aplicación "suficientemente aleatoria", entonces la esperanza de $T + L$ es $\sqrt{\pi N/2}$.*

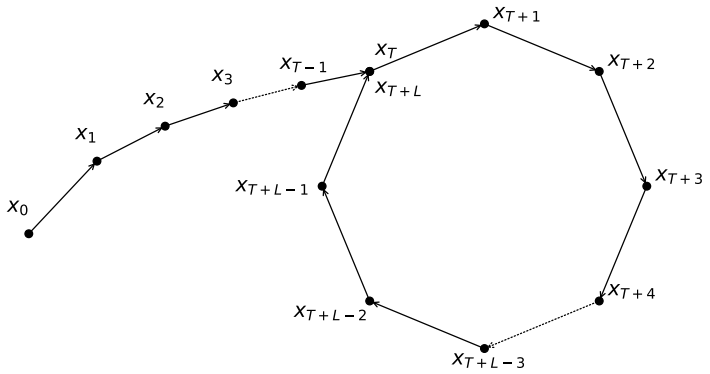


Figura II.1: Secuencia del algoritmo ρ de Pollard, cuya forma, similar a la de la letra griega ρ, da nombre al método.

Demostración. 1. De la definición de T y L sigue que, para dos índices i y j con $j > i$, se tiene que $x_j = x_i$ si y solo si $i \geq T$ y $j \equiv i \mod L$.

Por tanto, $x_{2i} = x_i$ si y solo si $i \geq T$ y $L \mid i$. El primer índice i que cumple esta condición debe estar en el intervalo $[T, T + L)$.

2. La probabilidad de que k puntos $x_0, x_1, \dots, x_{k-1}$ tomados al azar de S sean todos distintos es

$$\begin{aligned}\mathbb{P}\left(x_0, x_1, \dots, x_{k-1} \text{ son distintos}\right) &= \prod_{i=1}^{k-1} \mathbb{P}\left(x_i \neq x_j \ \ \forall\, 0 \leq j < i \,\middle|\, x_0, x_1, \dots, x_{i-1} \text{ son distintos}\right) \\ &= \prod_{i=1}^{k-1} \left(\frac{N-i}{N}\right) = \prod_{i=1}^{k-1} \left(1 - \frac{i}{N}\right).\end{aligned}$$

Para un t suficientemente pequeño, podemos emplear la aproximación $\log(1-t) \approx e^{-t}$. En nuestro caso, si suponemos que N es suficientemente grande y tomamos k del orden $\mathcal{O}(\sqrt{N})$, entonces i/N es pequeño para $i = 1, \dots, k-1$. Por consiguiente, podemos aproximar el producto anterior como

$$\mathbb{P}\left(x_0, x_1, \dots, x_{k-1} \text{ son distintos}\right) \approx \prod_{i=1}^{k-1} e^{\frac{-i}{N}} = e^{\frac{-1}{N}\sum_{i=1}^{k-1} i} = e^{-\frac{(k-1)k}{2N}} \approx e^{-\frac{k^2}{2N}}.$$

Por otro lado, si suponemos que $x_0, \dots, x_{k-1}$ son distintos, la probabilidad de que x_k sea igual a uno de ellos es k/N.

Teniendo en cuenta los dos resultados anteriores, obtenemos que

$$\begin{aligned}&\mathbb{P}(x_k \text{ es la primera repetición}) = \mathbb{P}\left(x_0, x_1, \dots, x_{k-1} \text{ son distintos y } x_k \text{ es una repetición}\right) \\ &\quad = \mathbb{P}\left(x_k \text{ es una repetición} \,\middle|\, x_0, x_1, \dots, x_{k-1} \text{ son distintos}\right) \mathbb{P}(x_0, x_1, \dots, x_{k-1} \text{ son distintos}) \\ &\quad \approx \frac{k}{N} e^{-\frac{k^2}{2N}}.\end{aligned}$$

Definiendo $\phi : \mathbb{R} \longrightarrow \mathbb{R}$ como $\phi(t) = t^2 e^{-\frac{t^2}{2}}$, concluimos que el valor esperado del número de iteraciones hasta que se encuentra la primera repetición es

$$\begin{aligned}\sum_{k\geq 1} k\ \mathbb{P}(x_k \text{ es la primera repetición}) &\approx \sum_{k\geq 1} \frac{k^2}{N} e^{-\frac{k^2}{2N}} = \sum_{k\geq 1} \phi(k/\sqrt{N}) \\ &\approx \sqrt{N}\int_0^\infty \phi(t)dt = \sqrt{N}\int_0^\infty t^2 e^{-\frac{t^2}{2}}dt \\ &= \sqrt{\frac{\pi N}{2}},\end{aligned}$$

usando que $\frac{1}{n}\sum_{k=1}^{\infty}\phi(k/n) \approx \int_0^\infty \phi(t)dt$. Para calcular la integral del último paso, basta considerar su cuadrado y aplicar un cambio de variables a coordenadas polares.

□

La idea del algoritmo de Pollard es tomar una función $f : (P) \longrightarrow (P)$ que tenga las características de una función aleatoria y que sea sencilla de calcular.

Para ello, se considera una partición aleatoria de (P) en subconjuntos $S_1, S_2, \ldots, S_M$ de aproximadamente el mismo tamaño. Habitualmente, se escoge $M = 16$ o $M = 32$. Por ejemplo, si $M = 32$, se puede imponer que un punto $X \in (P)$ se asigne a S_j si los cinco bits menos significativos de la coordenada x de X representan el entero $j-1$ en base 2.

Sea $H : (P) \longrightarrow \{1, 2, \ldots, M\}$ la *función de partición*, de forma que $H(P) = j$ si $P \in S_j$. Asimismo, sean a_j y b_j enteros elegidos uniformemente al azar en $\{0, 1, \ldots, M-1\}$ para cada $j = 1, \ldots, M$. Entonces, definimos la función $f : (P) \longrightarrow (P)$ como

$$f(X) = X + a_j P + b_j Q \quad \text{con } H(X) = j.$$

Nótese que si $X = cP + dQ$ para ciertos enteros $c, d \in [0, n-1]$, entonces $f(X) = \overline{c}P + \overline{d}Q$, donde $\overline{c} = c + a_j \pmod n$ y $\overline{d} = d + b_j \pmod n$.

Se construye entonces una secuencia de puntos $(X_i)_{i\geq 0}$ a partir de un punto inicial $X_0 \in (P)$. Para encontrar colisiones, se puede emplear el algoritmo de detección de ciclos de Floyd, que consiste en calcular pares de puntos de la forma (X_i, X_{2i}) para $i = 1, 2, \ldots$ hasta encontrar un índice para el que se cumpla que $X_i = X_{2i}$. Incluimos la especificación completa en el algoritmo 7.

Cabe destacar que la probabilidad de que el ataque falle (que en el paso 17 del algoritmo 7 resulte que $d' = d''$) es despreciable [4].

Se puede demostrar que la complejidad del ataque es $\mathcal{O}(\sqrt{n})$ [22]. Remarcamos también que el uso de memoria es muy pequeño, en contraposición con otros algoritmos de búsqueda de colisiones como *Babystep-Giantstep* [8].

Algoritmo 7 Algoritmo ρ de Pollard para el ECDLP.

Entrada: $P \in E(\mathbb{F}_q)$ de orden n primo, $Q \in (P)$.

Salida: $k = \log_P Q$.

Fijar el número de subconjuntos M.
Fijar una función de partición $H : (P) \longrightarrow \{1, 2, \dots, M\}$.
for $j = 1, \dots, M$ **do**
    Escoger $a_j, b_j \in \{0, 1, \dots, n-1\}$ uniformemente al azar.
    Calcular $R_j = a_j P + b_j Q$.
end for
Escoger $c', d' \in \{0, 1, \dots, n-1\}$ uniformemente al azar y calcular $X' = c'P + d'Q$.
Inicializar $X'' \leftarrow X'$, $c'' \leftarrow c'$ y $d'' \leftarrow d'$.
while $X' \neq X''$ **do**
    Calcular $j = H(X')$.
    Asignar $X' \leftarrow X' + R_j$, $c' \leftarrow c' + a_j \pmod n$, $d' \leftarrow d' + b_j \pmod n$.
    for i $= 1, 2$ **do**
        Calcular $j = H(X'')$.
        Asignar $X'' \leftarrow X'' + R_j$, $c'' \leftarrow c'' + a_j \pmod n$, $d'' \leftarrow d'' + b_j \pmod n$.
    end for
end while
if $d' = d''$ **then**
    Devolver: Ataque fallido.
else
    Devolver: $k = (c' - c'')(d'' - d')^{-1} \pmod n$.
end if

Bibliografía

Referencias principales

[1] Fulton, W. (1969). *Algebraic Curves. An introduction to algebraic geometry*, W. A. Benjamin, Inc., New York-Amsterdam.

[2] Garrity, T. et al. (2013). *Algebraic Geometry: A problem solving approach*, Student mathematical library: IAS/Park City mathematical subseries, **66**, American Mathematical Society.

[3] Gathman, A. (2023). *Plane Algebraic Curves* [Notas], RPTU Kaiserslautern. Disponible en https://agag-gathmann.math.rptu.de/class/curves-2023/curves-2023.pdf. Consultado por última vez el 15 de junio de 2025.

[4] Hankerson, D., Menezes A. y Vanstone S. (2004). *Guide to Elliptic Curve Cryptography*, 1st ed., Springer-Verlag, New York.

[5] Knapp, A. (1992). *Elliptic Curves*, Mathematical Notes, **40**, Princeton University Press.

[6] Kunz, E. (2005). *Introduction to Plane Algebraic Curves*, Birkhäuser Boston, MA.

[7] Milne, J. S. (2006). *Elliptic curves*, BookSurge Publishers.

[8] Silverman, J. H. (2009). *The arithmetic of elliptic curves*, 2nd ed., Graduate Texts in Mathematics, **106**, Springer-Verlag, New York.

[9] Silverman, J. H. y Tate, J. (2015). *Rational points on elliptic curves*, 2nd ed., Undergraduate Texts in Mathematics, Springer-Verlag, New York.

[10] Washington, L. (2008). *Elliptic Curves: Number Theory and Cryptography*, 2nd ed., Chapman and Hall/CRC.

Referencias complementarias

[11] Bosch, S. (2018). *Algebra: From the Viewpoint of Galois Theory*, Birkhäuser, Suiza.

[12] Cepelewicz, J. (2022). *'Post-Quantum' Cryptography Scheme Is Cracked on a Laptop*, Quanta Magazine [Online]. Disponible en https://www.quantamagazine.org/post-quantum-cryptography-scheme-is-cracked-on-a-laptop-20220824/. Consultado por última vez el 3 de mayo de 2025.

[13] Dujella, A. (2024). *History of elliptic curves rank records*, Universidad de Zagreb [Online]. Disponible en https://web.math.pmf.unizg.hr/~duje/tors/rankhist.html. Consultado por última vez el 20 de abril de 2025.

[14] Elkies, N. D. y Klagsbrun, Z. (2024). $\mathbb{Z}^{29}$ *in* $E(\mathbb{Q})$, Number Theory Listserver.

[15] Galbraith, S. D. (2001). Supersingular Curves in Cryptography, *Proceedings of the 7th International Conference on the Theory and Application of Cryptology and Information Security: Advances in Cryptology (ASIACRYPT '01)*, 495–513, Springer-Verlag, Berlin, Heidelberg.

[16] Glasscock, D. (2021). *Euler's elliptic integral addition theorem, repeated exponentiation, cotangent series and the Herglotz trick* [Notas], University of Massachusetts Lowell. Disponible en https://bpb-us-w2.wpmucdn.com/sites.uml.edu/dist/2/372/files/2021/06/eulerelliptic.pdf. Consultado por última vez el 8 de junio de 2025.

[17] Harding, D. et al. (2021). Transactions. *Bitcoin Developer Guide* [Online]. Disponible en https://developer.bitcoin.org/devguide/transactions.html. Consultado por última vez el 9 de febrero de 2025.

[18] Hartshorne, R. (1977). *Algebraic Geometry*, Graduate Texts in Mathematics, **52**, Springer-Verlag, New York.

[19] Hindry, M. y Silverman, J. H. (2000). *Diophantine Geometry: An Introduction*, Graduate Texts in Mathematics, **201**, Springer-Verlag, New York.

[20] Husemöller, D. (2004). *Elliptic Curves*, 2nd ed., Graduate Texts in Mathematics, **111**, Springer-Verlag, New York.

[21] Josefsson, S. y Liusvaara, I. (2017). *Edwards-Curve Digital Signature Algorithm (EdDSA)*, Internet Engineering Task Force, Request for Comments RFC 8032.

[22] Kim, J. H., Montenegro, R., Peres, Y. y Tetali, P. (2010). A Birthday Paradox for Markov chains with an optimal bound for collision in the Pollard Rho algorithm for discrete logarithm, *The Annals of Applied Probability*, **20**(2), 495-521.

[23] Koblitz, N. (1987). Elliptic curve cryptosystems, *Mathematics of Computation*, **48**(177), 203-209.

[24] Mazur, B. (1977). Modular curves and the Eisenstein ideal, *Publications Mathématiques de l'Institut des Hautes Études Scientifiques*, **47**, 33-186.

[25] McKernan, J. (2013). *12. Change of coordinates* [Notas], 203A Autumn 2013, UC San Diego. Disponible en https://mathweb.ucsd.edu/~jmckerna/Teaching/13-14/Autumn/203A. Consultado por última vez el 15 de abril de 2025.

[26] Merkle, J., y Lochter, M. (2010). *Elliptic Curve Cryptography (ECC) Brainpool Standard Curves and Curve Generation*, Internet Engineering Task Force, Request for Comments RFC 5639.

[27] Miller, V. (1986). Use of Elliptic Curves in Cryptography, *Advances in Cryptology - CRYPTO '85 Proceedings. Lecture Notes in Computer Science*, **85**, 417-426.

[28] Nakov, S. (2020). *Elliptic Curve Cryptography (ECC)* [Online], Practical Cryptography for Developers. Disponible en https://cryptobook.nakov.com/asymmetric-key-ciphers/elliptic-curve-cryptography-ecc. Consultado por última vez el 4 de mayo de 2025.

[29] National Institute of Standards and Technology (2023). *Digital Signature Standard (DSS)*, **186-5**, Federal Information Processing Standards Publication.

[30] Nir., Y. et al. (2018). *Elliptic Curve Cryptography (ECC) Cipher Suites for Transport Layer Security (TLS) Versions 1.2 and Earlier*, Internet Engineering Task Force (IETF), Request for Comments RFC 8422.

[31] Park, P. S. (2016). *Siegel's theorem over* $\mathbb{Q}$, Princeton University.

[32] Ravenel, D. (2007). *Elliptic Curves: what they are, why they are called elliptic, and why topologists like them, I* [Notas], Wayne State University. Disponible en https://people.math.rochester.edu/faculty/doug/mypapers/wayne1.pdf. Consultado por última vez el 15 de junio de 2025.

[33] Roy, B. y Vaishya, L. (2023). Elliptic curve over totally real fields: A Survey, *arXiv preprint arXiv:2304.09003*.

[34] Stallings, W. (2016). *Cryptography and Network Security: Principles and Practice*, 7nd ed., Pearson, Esssex.

[35] Standards for Efficient Cryptography Group (2010). *SEC 2: Recommended Elliptic Curve Domain Parameters*, Standards for Efficient Cryptography 2.

[36] Sutherland, A. (2017). *18.783 Elliptic Curves* [Notas], Massachusetts Institute of Technology. Disponible en https://dspace.mit.edu/handle/1721.1/122962. Consultado por última vez el 8 de junio de 2025.