10
Elliptic Curves

An elliptic curve is a nonsingular curve which can be defined by an equation the form $y^2 = x^3 + ax + b$ with $4a^3 + 27b^2 \neq 0$. Elliptic curves are one of the most exciting and central objects in modern number theory. Several very deep results were proved about elliptic curves in the last two decades, as we will see in Chapters 12, 13, and 14.[1] 1

This chapter establishes the basic foundations on which subsequent chapters will build. We begin with the definition of an elliptic curve over the complex numbers in Section 10.1. Section 10.5 is about how they have a natural structure of abelian group. In Section 10.6 we return to number theory, and consider the subgroup of points with coordinates in a fixed field, such as $\mathbb{Q}$.

Though we will use the word "natural" many times in this chapter, the reader should not try to assign too precise a meaning to this word.

10.1 The Definition of an Elliptic Curves

This section is about how to define elliptic curves over the complex number $\mathbb{C}$.

Recall from the introduction to this chapter that an elliptic is a curve that can be "defined" by an equation of the form $y^2 = x^3 + ax + b$. This is technically correct, but sneaky, because the phrase "can be defined" is

[1] Expand and elaborate.

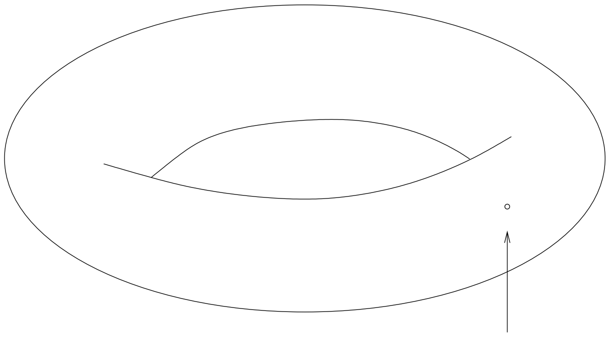

FIGURE 10.1. Y is Homeomorphic to a Torus With a Point Removed

vague. The graph Y of $y^2 = x^3 + ax + b$ is the set of points in $\mathbb{C}^2 = \mathbb{C} \times \mathbb{C}$ that satisfy $y^2 = x^3 + ax + b$, so

$$Y = \{(z, w) \in \mathbb{C}^2 : w^2 = z^3 + az + b\} \subset \mathbb{C}^2.$$

Since for each value of z there are exactly two values of w that satisfy the equation (counting "multiplicity"), Y has dimension 2 when viewed as a real manifold (locally Y looks like a small open subset of $\mathbb{C}$). When viewed as a complex manifold, Y has dimension 1, which is one reason that we refer to Y as a "curve".

Using the Weierstrass $\wp$ function from complex analysis (which will appear again in Section 10.3.8) one can see that there is a homeomorphism between Y and a torus with one point removed (see Figure 10.1). Notice that Y is "incomplete" in the sense that it is visibly missing a point. The set Y is closed when viewed as a subset of $\mathbb{C}^2$, since it is the inverse image of 0 under the continuous map $\mathbb{C}^2 \to \mathbb{C}$ given by $(x, y) \mapsto y^2 - x^3 - ax - b$. Thus we won't find the missing point by taking the closure of Y in $\mathbb{C}^2$. We instead consider Y as a subset of the project plane, where the closure will include the one missing point.

10.1.1 Some Topology

Before defining the projective plane, we review some basic topology.

A *topological space* is a set X together with a collection of open subsets $U \subset X$ that are closed under arbitrary unions, finite intersections, and X and $\emptyset$ are open. A subset $A \subset X$ is *closed* if $X \setminus A$ is open.

A *continuous map* $f : X \to Y$ of topological spaces is a map such that whenever $U \subset Y$ is open, the inverse image $f^{-1}(U)$ is open in X. Because $f^{-1}(X \setminus A) = X \setminus f^{-1}(A)$, one sees at once that f is continuous if and only if the inverse image of every closed subset of Y is closed in X.

A subset $U \subset X$ is *clopen* if it is both open and closed. A topological space X is *connected* if it has no clopen subsets besides $\emptyset$ and X. The continuous image of a connected set is connected, because if $f : X \to Y$ is continuous and $C \subset Y$ is clopen, then $f^{-1}(C) \subset X$ is also clopen, and $f^{-1}(C) = X$ or $\emptyset$ if and only if $C = Y$ or $\emptyset$.

Suppose X is a topological space and $f : X \to Y$ is a map from X to a set Y. The *induced topology* on Y is the topology in which the open subsets of Y are the subsets $U \subset Y$ such that $f^{-1}(U)$ is an open subset of X. This is the "coarsest" topology on Y that makes the map f continuous.

We assume the reader is familiar with the standard topology on $\mathbb{R}$ and $\mathbb{C}$ from a course in Calculus or Analysis (open subsets of $\mathbb{R}$ are unions of open intervals (a, b), etc.).

10.1.2 The Projective Plane

Definition 10.1.1 (Projective Plane). The *projective plane* $\mathbb{P}^2$ is the set of triples $(a, b, c) \in \mathbb{C}^3$ with a, b, c not all 0, modulo the equivalence relation

$$(a, b, c) \sim (\lambda a, \lambda b, \lambda c)$$

for all nonzero $\lambda \in \mathbb{C}$. Denote by $(a : b : c)$ the equivalence class of (a, b, c). The topology on $\mathbb{P}^2$ is the one induced by viewing it as a quotient of $\mathbb{C}^3 \setminus \{0\}$.

The projective plane is a little bigger than the usual plane, in the following sense. There is a map $\mathbb{C}^2 \hookrightarrow \mathbb{P}^2$ that sends (a, b) to $(a : b : 1)$, and the complement of the image is a projective line:

$$\mathbb{P}^2 \setminus \mathbb{C}^2 \cong \{(a : 1 : 0) : a \in \mathbb{C}\} \cup \{(1 : 0 : 0)\}.$$

Thus $\mathbb{P}^2$ is set-theoretically the disjoint union

$$\mathbb{C}^2 \cup \mathbb{C} \cup \{\text{point}\}.$$

Since $\mathbb{P}^2$ is a continuous image of the connected set $\mathbb{C}^3 \setminus \{0\}$, we see that $\mathbb{P}^2$ is also connected, so we should not view $\mathbb{P}^2$ "topologically" as the above disjoint union. The inverse image of $\mathbb{C}^2$ in $\mathbb{C}^3 \setminus \{0\}$ is the complement of the (closed) plane $\{(a, b, 0) : a, b \in \mathbb{C}\}$, so $\mathbb{C}^2$ is an open subset of $\mathbb{P}^2$. Since $\mathbb{P}^2$ is connected, $\mathbb{C}^2$ is *not* a closed subset of $\mathbb{P}^2$.

Exploring this idea further, it is useful to view $\mathbb{P}^2$ as being covered by 3 copies of $\mathbb{C}^2$, though not disjointly. We consider three ways to embed $\mathbb{C}^2$ as an open subset of $\mathbb{P}^2$; these three embeddings send (a, b) to $(1 \colon a \colon b)$, $(a \colon 1 \colon b)$, and $(a \colon b \colon 1)$, respectively. We will denote the three images of $\mathbb{C}^2$ by U_1, U_2, and U_3, respectively. Notice that $\mathbb{P}^2 = U_1 \cup U_2 \cup U_3$, but that the union is not disjoint. In order to "see" a subset S of $\mathbb{P}^2$, it is often useful to look at $S \cap U_i$ for each i. For example, let's compute $Y \cap U_2$.

Lemma 10.1.2. *We have*

$$Y \cap U_2 = \left\{ \left(\frac{x}{y} : 1 : \frac{1}{y} \right) : y^2 = x^3 + ax + b \text{ and } y \neq 0 \right\}.$$

In particular, Y is not closed in $\mathbb{P}^2$, since $(0\colon 1\colon 0)$ is a limit point of $Y \cap U_2$ that is not contained in $Y \cap U_2$.

Proof. The first equality follows easily from the definitions. To see that $(0\colon 1\colon 0)$ is a limit point, let $\alpha^{1/2}$ denote the positive square root of the positive real number α. Then $|x/y| = |x|/|x^3 + ax + b|^{1/2} \to 0$ and $|1/y| = 1/|x^3 + ax + b|^{1/2} \to 0$ as $|x| \to \infty$. □

10.1.3 The Closure of $y^2 = x^3 + ax + b$ in $\mathbb{P}^2$

Proposition 10.1.3. *The closure of the graph Y of $y^2 = x^3 + ax + b$ in $\mathbb{P}^2$ is the graph X of $y^2z = x^3 + axz^2 + bz^3$ in $\mathbb{P}^2$. We have*

$$\begin{aligned} X &= \{(x : y : z) \in \mathbb{P}^2 : y^2 z = x^3 + axz^2 + bz^3\} \\ &= Y \cup \{(0 : 1 : 0)\}. \end{aligned}$$

Proof. The inverse of image of X in $\mathbb{C}^3 \setminus \{0\}$ is the inverse image of 0 under the continuous map $\mathbb{C}^3 \setminus \{0\} \to \mathbb{C}$ defined by $(x, y, z) \mapsto y^2z - (x^3 + axz^2 + bz^3)$, so X is closed. The difference $X \setminus Y$ is the set of points $(x : y : 0)$ that satisfy $y^2z = x^3 + axz^2 + bz^3$, so $X \setminus Y = \{(0 : 1 : 0)\}$. Thus X is closed and $X = Y \cup \{(0 : 1 : 0)\}$. By Lemma 10.1.2, Y is not closed in $\mathbb{P}^2$, so X is the closure of Y in $\mathbb{P}^2$. □

We will frequently refer to the point $(0\colon 1\colon 0)$ on X as the "point at infinity".

Definition 10.1.4 (Elliptic Curve). An *elliptic curve* E over the complex numbers $\mathbb{C}$ is a closed cubic curve in $\mathbb{P}^2$ defined by an equation

$$y^2 = x^3 + ax + b,$$

with $a, b \in \mathbb{C}$ and $\Delta = -16(4a^3 + 27b^2) \neq 0$. We let $E(\mathbb{C})$ denote the underlying set of points on E in $\mathbb{P}^2$.

If $4a^3 + 27b^2 = 0$ then the cubic $x^3 + ax + b$ has a repeated root α. Locally at $(\alpha, 0)$, Y is does not behave like an open subset of $\mathbb{C}$, which is why we exclude this case.

Remark 10.1.5. An ellipse is the graph of $ax^2 + by^2 = r$ with $a, b, r > 0$. Elliptic curves are not ellipses; they are called "elliptic" because they arise when studying arc lengths of ellipses (see Exercise 7). They haven't always been called elliptic curves; e.g., in the early 1960s Cassels called them "abelian varieties of dimension one" (see [Cas62]).