CHAPTER I

From Congruent Numbers to Elliptic Curves

The theory of elliptic curves and modular forms is one subject where the most diverse branches of mathematics come together: complex analysis, algebraic geometry, representation theory, number theory. While our point of view will be number theoretic, we shall find ourselves using the type of techniques that one learns in basic courses in complex variables, real variables, and algebra. A well-known feature of number theory is the abundance of conjectures and theorems whose statements are accessible to high school students but whose proofs either are unknown or, in some cases, are the culmination of decades of research using some of the most powerful tools of twentieth century mathematics.

We shall motivate our choice of topics by one such theorem: an elegant characterization of so-called "congruent numbers" that was recently proved by J. Tunnell [Tunnell 1983]. A few of the proofs of necessary results go beyond our scope, but many of the ingredients in the proof of Tunnell's theorem will be developed in complete detail.

Tunnell's theorem gives an almost complete answer to an ancient problem: find a simple test to determine whether or not a given integer n is the area of some right triangle all of whose sides are rational numbers. A natural number n is called "congruent" if there exists a right triangle with all three sides rational and area n. For example, 6 is the area of the 3–4–5 right triangle, and so is a congruent number.

Right triangles whose sides are integers X, Y, Z (a "Pythagorean triple") were studied in ancient Greece by Pythagoras, Euclid, Diophantus, and others. Their central discovery was that there is an easy way to generate all such triangles. Namely, take any two positive integers a and b with $a > b$, draw the line in the uv-plane through the point $(-1, 0)$ with slope b/a. Let (u, v) be the second point of intersection of this line with the unit circle (see Fig. I.1). It is not hard to show that

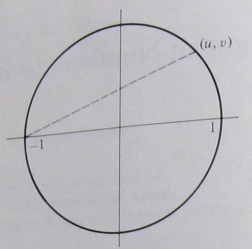

Figure I.1

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

Then the integers $X = a^2 - b^2$, $Y = 2ab$, $Z = a^2 + b^2$ are the sides of a right triangle; the fact that $X^2 + Y^2 = Z^2$ follows because $u^2 + v^2 = 1$. By letting a and b range through all positive integers with $a > b$, one gets all possible Pythagorean triples (see Problem 1 below).

Although the problem of studying numbers n which occur as areas of rational right triangles was of interest to the Greeks in special cases, it seems that the congruent number problem was first discussed systematically by Arab scholars of the tenth century. (For a detailed history of the problem of determining which numbers are "congruent", see [L. E. Dickson 1952, Ch. XVI]; see also [Guy 1981, Section D27].) The Arab investigators preferred to rephrase the problem in the following equivalent form: given n, can one find a rational number x such that $x^2 + n$ and $x^2 - n$ are both squares of rational numbers? (The equivalence of these two forms of the congruent number problem was known to the Greeks and to the Arabs; for a proof of this elementary fact, see Proposition 1 below.)

Since that time, some well-known mathematicians have devoted considerable energy to special cases of the congruent number problem. For example, Euler was the first to show that $n = 7$ is a congruent number. Fermat showed that $n = 1$ is *not*; this result is essentially equivalent to Fermat's Last Theorem for the exponent 4 (i.e., the fact that $X^4 + Y^4 = Z^4$ has no nontrivial integer solutions).

It eventually became known that the numbers 1, 2, 3, 4 are not congruent numbers, but 5, 6, 7 are. However, it looked hopeless to find a straightforward criterion to tell whether or not a given n is congruent. A major advance in the twentieth century was to place this problem in the context of the arithmetic theory of elliptic curves. It was in this context that Tunnell was able to prove his remarkable theorem.

Here is part of what Tunnell's theorem says (the full statement will be given later):

Theorem (Tunnell). *Let n be an odd squarefree natural number. Consider the two conditions:*

(A) *n is congruent;*
(B) *the number of triples of integers (x, y, z) satisfying $2x^2 + y^2 + 8z^2 = n$ is equal to twice the number of triples satisfying $2x^2 + y^2 + 32z^2 = n$.*

Then (A) *implies* (B); *and, if a weak form of the so-called Birch–Swinnerton-Dyer conjecture is true, then* (B) *also implies* (A).

The central concepts in the proof of Tunnell's theorem—the Hasse–Weil L-function of an elliptic curve, the Birch–Swinnerton-Dyer conjecture, modular forms of half integer weight—will be discussed in later chapters. Our concern in this chapter will be to establish the connection between congruent numbers and a certain family of elliptic curves, in the process giving the definition and some basic properties of elliptic curves.

§1. Congruent numbers

Let us first make a more general definition of a congruent number. A positive rational number $r \in \mathbb{Q}$ is called a "congruent number" if it is the area of some right triangle with rational sides. Suppose r is congruent, and $X, Y, Z \in \mathbb{Q}$ are the sides of a triangle with area r. For any nonzero $r \in \mathbb{Q}$ we can find some $s \in \mathbb{Q}$ such that $s^2 r$ is a squarefree integer. But the triangle with sides sX, sY, sZ has area $s^2 r$. Thus, without loss of generality we may assume that $r = n$ is a squarefree natural number. Expressed in group language, we can say that whether or not a number r in the multiplicative group $\mathbb{Q}^+$ of positive rational numbers has the congruent property depends only on its coset modulo the subgroup $(\mathbb{Q}^+)^2$ consisting of the squares of rational numbers; and each coset in $\mathbb{Q}^+/(\mathbb{Q}^+)^2$ contains a unique squarefree natural number $r = n$. In what follows, when speaking of congruent numbers, we shall always assume that the number is a squarefree positive integer.

Notice that the definition of a congruent number does not require the sides of the triangle to be integral, only rational. While $n = 6$ is the smallest possible area of a right triangle with integer sides, one can find right triangles with rational sides having area $n = 5$. The right triangle with sides $1\frac{1}{2}$, $6\frac{2}{3}$, $6\frac{2}{6}$ is such a triangle (see Fig. I.2). It turns out that $n = 5$ is the smallest congruent number (recall that we are using "congruent number" to mean "congruent squarefree natural number").

There is a simple algorithm using Pythagorean triples (see the problems below) that will eventually list all congruent numbers. Unfortunately, given

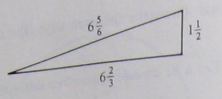

Figure I.2

n, one cannot tell how long one must wait to get n if it is congruent; thus, if n has not appeared we do not know whether this means that n is not a congruent number or that we have simply not waited long enough. From a practical point of view, the beauty of Tunnell's theorem is that his condition (B) can be easily and rapidly verified by an effective algorithm. Thus, his theorem almost settles the congruent number problem, i.e., the problem of finding a verifiable criterion for whether a given n is congruent. We must say "almost settles" because in one direction the criterion is only known to work in all cases if one assumes a conjecture about elliptic curves.

Now suppose that X, Y, Z are the sides of a right triangle with area n. This means: $X^2 + Y^2 = Z^2$, and $\frac{1}{2}XY = n$. Thus, algebraically speaking, the condition that n be a congruent number says that these two equations have a simultaneous solution $X, Y, Z \in \mathbb{Q}$. In the proposition that follows, we derive an alternate condition for n to be a congruent number. In listing triangles with sides X, Y, Z, we shall not want to list X, Y, Z and Y, X, Z separately. So for now let us fix the ordering by requiring that $X < Y < Z$ (Z is the hypotenuse).

Proposition 1. *Let n be a fixed squarefree positive integer. Let X, Y, Z, x always denote rational numbers, with $X < Y < Z$. There is a one-to-one correspondence between right triangles with legs X and Y, hypotenuse Z, and area n; and numbers x for which x, $x+n$, and $x-n$ are each the square of a rational number. The correspondence is:*

$$X, Y, Z \to x = (Z/2)^2$$

$$x \to X = \sqrt{x+n} - \sqrt{x-n}, \quad Y = \sqrt{x+n} + \sqrt{x-n}, \quad Z = 2\sqrt{x}.$$

In particular, n is a congruent number if and only if there exists x such that x, $x+n$, and $x-n$ are squares of rational numbers.

PROOF. First suppose that X, Y, Z is a triple with the desired properties: $X^2 + Y^2 = Z^2$, $\frac{1}{2}XY = n$. If we add or subtract four times the second equation from the first, we obtain: $(X \pm Y)^2 = Z^2 \pm 4n$. If we then divide both sides by four, we see that $x = (Z/2)^2$ has the property that the numbers $x \pm n$ are the squares of $(X \pm Y)/2$. Conversely, given x with the desired properties, it is easy to see that the three positive rational numbers $X < Y < Z$ given by the formulas in the proposition satisfy: $XY = 2n$, and $X^2 + Y^2 = 4x = Z^2$. Finally, to establish the one-to-one correspondence, it only remains

to verify that no two distinct triples X, Y, Z can lead to the same x. We leave this to the reader (see the problems below). □

PROBLEMS

1. Recall that a Pythagorean triple is a solution (X, Y, Z) in positive integers to the equation $X^2 + Y^2 = Z^2$. It is called "primitive" if X, Y, Z have no common factor. Suppose that $a > b$ are two relatively prime positive integers, not both odd. Show that $X = a^2 - b^2$, $Y = 2ab$, $Z = a^2 + b^2$ form a primitive Pythagorean triple, and that all primitive Pythagorean triples are obtained in this way.

2. Use Problem 1 to write a flowchart for an algorithm that lists all squarefree congruent numbers (of course, not in increasing order). List the first twelve distinct congruent numbers your algorithm gives. Note that there is no way of knowing when a given congruent number n will appear in the list. For example, 101 is a congruent number, but the first Pythagorean triple which leads to an area s^2 101 involves twenty-two-digit numbers (see [Guy 1981, p. 106]). One hundred fifty-seven is even worse (see Fig. I.3). One cannot use this algorithm to establish that some n is *not* a congruent number. Technically, it is not a real algorithm, only a "semi-algorithm".

3. (a) Show that if 1 were a congruent number, then the equation $x^4 - y^4 = u^2$ would have an integer solution with u odd.
   (b) Prove that 1 is not a congruent number. (Note: A consequence is Fermat's Last Theorem for the exponent 4.)

4. Finish the proof of Proposition 1 by showing that no two triples X, Y, Z can lead to the same x.

5. (a) Find $x \in (\mathbb{Q}^+)^2$ such that $x \pm 5 \in (\mathbb{Q}^+)^2$.
   (b) Find $x \in (\mathbb{Q}^+)^2$ such that $x \pm 6 \in (\mathbb{Q}^+)^2$.

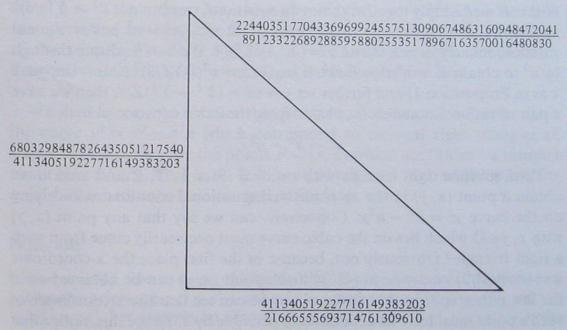

Figure I.3. The Simplest Rational Right Triangle with Area 157 (computed by D. Zagier).

(c) Find two values $x \in (\mathbb{Q}^+)^2$ such that $x \pm 210 \in (\mathbb{Q}^+)^2$. At the end of this chapter we shall prove that if there is one such x, then there are infinitely many. Equivalently (by Proposition 1), if there exists one right triangle with rational sides and area n, then there exist infinitely many.

6. (a) Show that condition (B) in Tunnell's theorem is equivalent to the condition that the number of ways n can be written in the form $2x^2 + y^2 + 8z^2$ with x, y, z integers and z odd, be equal to the number of ways n can be written in this form with z even.
   (b) Write a flowchart for an algorithm that tests condition (B) in Tunnell's theorem for a given n.

7. (a) Prove that condition (B) in Tunnell's theorem always holds if n is congruent to 5 or 7 modulo 8.
   (b) Check condition (B) for all squarefree $n \equiv 1$ or 3 (mod 8) until you find such an n for which condition (B) holds.
   (c) By Tunnell's theorem, the number you found in part (b) should be the smallest congruent number congruent to 1 or 3 modulo 8. Use the algorithm in Problem 2 to find a right triangle with rational sides and area equal to the number you found in part (b).

§2. A certain cubic equation

In this section we find yet another equivalent characterization of congruent numbers.

In the proof of Proposition 1 in the last section, we arrived at the equations $((X \pm Y)/2)^2 = (Z/2)^2 \pm n$ whenever X, Y, Z are the sides of a triangle with area n. If we multiply together these two equations, we obtain $((X^2 - Y^2)/4)^2 = (Z/2)^4 - n^2$. This shows that the equation $u^4 - n^2 = v^2$ has a rational solution, namely, $u = Z/2$ and $v = (X^2 - Y^2)/4$. We next multiply through by u^2 to obtain $u^6 - n^2u^2 = (uv)^2$. If we set $x = u^2 = (Z/2)^2$ (this is the same x as in Proposition 1) and further set $y = uv = (X^2 - Y^2)Z/8$, then we have a pair of rational numbers (x, y) satisfying the cubic equation:

$$y^2 = x^3 - n^2x.$$

Thus, given a right triangle with rational sides X, Y, Z and area n, we obtain a point (x, y) in the xy-plane having rational coordinates and lying on the curve $y^2 = x^3 - n^2x$. Conversely, can we say that any point (x, y) with $x, y \in \mathbb{Q}$ which lies on the cubic curve must necessarily come from such a right triangle? Obviously not, because in the first place the x-coordinate $x = u^2 = (Z/2)^2$ must lie in $(\mathbb{Q}^+)^2$ if the point (x, y) can be obtained as in the last paragraph. In the second place, we can see that the x-coordinate of such a point must have its denominator divisible by 2. To see this, notice that the triangle X, Y, Z can be obtained starting with a primitive Pythagorean triple X', Y', Z' corresponding to a right triangle with integral sides X', Y', Z' and area s^2n, and then dividing the sides by s to get X, Y, Z. But in a primitive

Pythagorean triple X' and Y' have different parity, and Z' is odd. We conclude that (1) $x = (Z/2)^2 = (Z'/2s)^2$ has denominator divisible by 2 and (2) the power of 2 dividing the denominator of Z is equal to the power of 2 dividing the denominator of one of the other two sides, while a strictly lower power of 2 divides the denominator of the third side. (For example, in the triangle in Fig. I.2 with area 5, the hypotenuse and the shorter side have a 2 in the denominator, while the other leg does not.) We conclude that a *necessary* condition for the point (x, y) with rational coordinates on the curve $y^2 = x^3 - n^2x$ to come from a right triangle is that x be a square and that its denominator be divisible by 2. For example, when $n = 31$, the point $(41^2/7^2, 29520/7^3)$ on the curve $y^2 = x^3 - 31^2x$ does not come from a triangle, even though its x-coordinate is a square. We next prove that these two conditions are *sufficient* for a point on the curve to come from a triangle.

Proposition 2. *Let (x, y) be a point with rational coordinates on the curve $y^2 = x^3 - n^2x$. Suppose that x satisfies the two conditions:* (i) *it is the square of a rational number and* (ii) *its denominator is even. Then there exists a right triangle with rational sides and area n which corresponds to x under the correspondence in Proposition* 1.

PROOF. Let $u = \sqrt{x} \in \mathbb{Q}^+$. We work backwards through the sequence of steps at the beginning of this section. That is, set $v = y/u$, so that $v^2 = y^2/x = x^2 - n^2$, i.e., $v^2 + n^2 = x^2$. Now let t be the denominator of u, i.e., the smallest positive integer such that $tu \in \mathbb{Z}$. By assumption, t is even. Notice that the denominators of v^2 and x^2 are the same (because n is an integer, and $v^2 + n^2 = x^2$), and this denominator is t^4. Thus, t^2v, t^2n, t^2x is a primitive Pythagorean triple, with t^2n even. By Problem 1 of §1, there exist integers a and b such that: $t^2n = 2ab$, $t^2v = a^2 - b^2$, $t^2x = a^2 + b^2$. Then the right triangle with sides $2a/t$, $2b/t$, $2u$ has area $2ab/t^2 = n$, as desired. The image of this triangle $X = 2a/t$, $Y = 2b/t$, $Z = 2u$ under the correspondence in Proposition 1 is $x = (Z/2)^2 = u^2$. This proves Proposition 2. □

We shall later prove another characterization of the points $P = (x, y)$ on the curve $y^2 = x^3 - n^2x$ which correspond to rational right triangles of area n. Namely, they are the points $P = (x, y)$ which are "twice" a rational point $P' = (x', y')$. That is, $P' + P' = P$, where "+" is an addition law for points on our curve, which we shall define later.

PROBLEMS

1. Find a simple linear change of variables that gives a one-to-one correspondence between points on $ny^2 = x^3 + ax^2 + bx + c$ and points on $y^2 = x^3 + anx^2 + bn^2x + cn^3$. For example, an alternate form of the equation $y^2 = x^3 - n^2x$ is the equation $ny^2 = x^3 - x$.

2. Another correspondence between rational right triangles X, Y, Z with area $\frac{1}{2}XY = n$ and rational solutions to $y^2 = x^3 - n^2x$ can be constructed as follows.

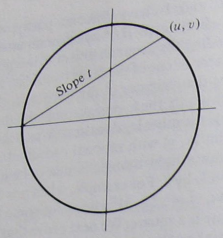

Figure I.4

(a) Parametrize all right triangles by letting the point $u = X/Z$, $v = Y/Z$ on the unit circle correspond to the slope t of the line joining $(-1, 0)$ to this point (see Fig. I.4). Show that

$$u = \frac{1-t^2}{1+t^2}, \qquad v = \frac{2t}{1+t^2}.$$

(Note: This is the usual way to parametrize a conic. If $t = a/b$ is rational, then the point (u, v) corresponds to the Pythagorean triple constructed by the method at the beginning of the chapter.)

(b) If we want the triangle X, Y, Z to have area n, express n/Z^2 in terms of t.

(c) Show that the point $x = -nt$, $y = n^2(1+t^2)/Z$ is on the curve $y^2 = x^3 - n^2x$. Express (x, y) in terms of X, Y, Z.

(d) Conversely, show that any point (x, y) on the curve $y^2 = x^3 - n^2x$ with $y \neq 0$ comes from a triangle, except that to get points with positive x, we must allow triangles with negative X and Y (but positive area $\frac{1}{2}XY = n$), and to get points with negative y we must allow negative Z (see Fig. I.5). Later in this chapter we shall show the connection between this correspondence and the one given in the text above.

(e) Find the points on $y^2 = x^3 - 36x$ coming from the 3–4–5 right triangle and all equivalent triangles (4–3–5, (−3)–(−4)–5, etc.).

3. Generalize the congruent number problem as follows. Fix an angle θ not necessarily 90°. But suppose that $A = \cos\theta$ and $B = \sin\theta$ are both rational. Let n be a square-free natural number. One can then ask whether n is the area of any triangle with rational sides one of whose angles is θ.

(a) Show that the answer to this question is equivalent to a question about rational solutions to a certain cubic equation (whose coefficients depend on θ as well as n).

(b) Suppose that the line joining the point $(-1, 0)$ to the point (A, B) on the unit circle has slope λ. Show that the cubic in part (a) is equivalent (by a linear change of variables) to the cubic $ny^2 = x(x - \lambda)(x + (1/\lambda))$. The classical congruent number problem is, of course, the case $\lambda = 1$.

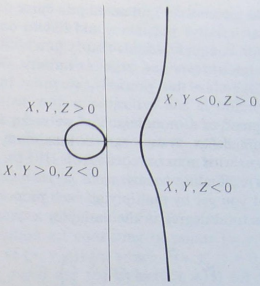

Figure I.5

§3. Elliptic curves

The locus of points $P = (x, y)$ satisfying $y^2 = x^3 - n^2x$ is a special case of what's called an "elliptic curve". More generally, let K be any field, and let $f(x) \in K[x]$ be a cubic polynomial with coefficients in K which has *distinct* roots (perhaps in some extension of K). We shall suppose that K does *not* have characteristic 2. Then the solutions to the equation

$$y^2 = f(x), \tag{3.1}$$

where x and y are in some extension K' of K, are called the *K'-points of the elliptic curve defined by* (3.1). We have just been dealing with the example $K = K' = \mathbb{Q}$, $f(x) = x^3 - n^2x$. Note that this example $y^2 = x^3 - n^2x$ satisfies the condition for an elliptic curve over any field K of characteristic p, as long as p does not divide $2n$, since the three roots $0, \pm n$ of $f(x) = x^3 - n^2x$ are then distinct.

In general, if $x_0, y_0 \in K'$ are the coordinates of a point on a curve C defined by an equation $F(x, y) = 0$, we say that C is "smooth" at (x_0, y_0) if the two partial derivatives $\partial F/\partial x$ and $\partial F/\partial y$ are not both zero at (x_0, y_0). This is the definition regardless of the ground field (the partial derivative of a polynomial $F(x, y)$ is defined by the usual formula, which makes sense over any field). If K' is the field $\mathbb{R}$ of real numbers, this agrees with the usual condition for C to have a tangent line. In the case $F(x, y) = y^2 - f(x)$, the partial derivatives are $2y_0$ and $-f'(x_0)$. Since K' is not a field of characteristic 2, these vanish simultaneously if and only if $y_0 = 0$ and x_0 is a multiple root of $f(x)$. Thus, the curve has a non-smooth point if and only if $f(x)$ has a multiple root. It is for this reason that we assumed distinct roots in the definition of an elliptic curve: an elliptic curve is smooth at all of its points.

In addition to the points (x, y) on an elliptic curve (3.1), there is a very important "point at infinity" that we would like to consider as being on the curve, much as in complex variable theory in addition to the points on the complex plane one throws in a point at infinity, thereby forming the "Riemann sphere". To do this precisely, we now introduce projective coordinates.

By the "total degree" of a monomial x^iy^j we mean $i + j$. By the "total degree" of a polynomial $F(x, y)$ we mean the maximum total degree of the monomials that occur with nonzero coefficients. If $F(x, y)$ has total degree n, we define the corresponding *homogeneous polynomial* $\tilde{F}(x, y, z)$ of *three* variables to be what you get by multiplying each monomial x^iy^j in $F(x, y)$ by z^{n-i-j} to bring its total degree in the variables x, y, z up to n; in other words,

$$\tilde{F}(x, y, z) = z^nF\left(\frac{x}{z}, \frac{y}{z}\right).$$

In our example $F(x, y) = y^2 - (x^3 - n^2x)$, we have $\tilde{F}(x, y, z) = y^2z - x^3 + n^2xz^2$. Notice that $F(x, y) = \tilde{F}(x, y, 1)$.

Suppose that our polynomials have coefficients in a field K, and we are interested in triples $x, y, z \in K$ such that $\tilde{F}(x, y, z) = 0$. Notice that:

(1) for any $\lambda \in K$, $\tilde{F}(\lambda x, \lambda y, \lambda z) = \lambda^n\tilde{F}(x, y, z)$ (n = total degree of F);

(2) for any nonzero $\lambda \in K$, $\tilde{F}(\lambda x, \lambda y, \lambda z) = 0$ if and only if $\tilde{F}(x, y, z) = 0$. In particular, for $z \neq 0$ we have $\tilde{F}(x, y, z) = 0$ if and only if $F(x/z, y/z) = 0$.

Because of (2), it is natural to look at equivalence classes of triples $x, y, z \in K$, where we say that two triples (x, y, z) and (x', y', z') are equivalent if there exists a nonzero $\lambda \in K$ such that $(x', y', z') = \lambda(x, y, z)$. We omit the trivial triple $(0, 0, 0)$, and then we define the "projective plane $\mathbb{P}^2_K$" to be the set of all equivalence classes of nontrivial triples.

No normal person likes to think in terms of "equivalence classes", and fortunately there are more visual ways to think of the projective plane. Suppose that K is the field $\mathbb{R}$ of real numbers. Then the triples (x, y, z) in an equivalence class all correspond to points in three-dimensional Euclidean space lying on a line through the origin. Thus, $\mathbb{P}^2_{\mathbb{R}}$ can be thought of geometrically as the set of lines through the origin in three-dimensional space.

Another way to visualize $\mathbb{P}^2_{\mathbb{R}}$ is to place a plane at a distance from the origin in three-dimensional space, for example, take the plane parallel to the xy-plane and at a distance 1 from it, i.e., the plane with equation $z = 1$. All lines through the origin, except for those lying in the xy-plane, have a unique point of intersection with this plane. That is, every equivalence class of triples (x, y, z) with nonzero z-coordinate has a unique triple of the form $(x, y, 1)$. So we think of such equivalence classes as points in the ordinary xy-plane. The remaining triples, those of the form $(x, y, 0)$, make up the "line at infinity".

The line at infinity, in turn, can be visualized as an ordinary line (say, the line $y = 1$ in the xy-plane) consisting of the equivalence classes with nonzero y-coordinate and hence containing a unique triple of the form $(x, 1, 0)$, together with a single "point at infinity" $(1, 0, 0)$. That is, we define the projective line $\mathbb{P}^1_K$ over a field K to be the set of equivalence classes of pairs (x, y) with $(x, y) \sim (\lambda x, \lambda y)$. Then $\mathbb{P}^2_K$ can be thought of as an ordinary plane $(x, y, 1)$ together with a projective line at infinity, which, in turn, consists of an ordinary line $(x, 1, 0)$ together with its point at infinity $(1, 0, 0)$.

More generally, n-dimensional projective space $\mathbb{P}^n_K$ is defined using equivalence classes of $(n + 1)$-tuples, and can be visualized as the usual space of n-tuples $(x_1, \ldots, x_n, 1)$ together with a $\mathbb{P}^{n-1}_K$ at infinity. But we shall only have need of $\mathbb{P}^1_K$ and $\mathbb{P}^2_K$.

Given a homogeneous polynomial $\tilde{F}(x, y, z)$ with coefficients in K, we can look at the solution set consisting of points (x, y, z) in $\mathbb{P}^2_K$ (actually, equivalence classes of (x, y, z)) for which $\tilde{F}(x, y, z) = 0$. The points of this solution set where $z \neq 0$ are the points $(x, y, 1)$ for which $\tilde{F}(x, y, 1) = F(x, y) = 0$. The remaining points are on the line at infinity. The solution set of $\tilde{F}(x, y, z) = 0$ is called the "projective completion" of the curve $F(x, y) = 0$. From now on, when we speak of a "line", a "conic section", an "elliptic curve", etc., we shall usually be working in a projective plane $\mathbb{P}^2_K$, in which case these terms will always denote the projective completion of the usual curve in the xy-plane. For example, the line $y = mx + b$ will really mean the solution set to $y = mx + bz$ in $\mathbb{P}^2_K$; and the elliptic curve $y^2 = x^3 - n^2x$ will now mean the solution set to $y^2z = x^3 - n^2xz^2$ in $\mathbb{P}^2_K$.

Let us look more closely at our favorite example: $F(x, y) = y^2 - x^3 + n^2x$, $\tilde{F}(x, y, z) = y^2z - x^3 + n^2xz^2$. The points at infinity on this elliptic curve are the equivalence classes $(x, y, 0)$ such that $0 = \tilde{F}(x, y, 0) = -x^3$. i.e., $x = 0$. There is only one such equivalence class $(0, 1, 0)$. Intuitively, if we take $K = \mathbb{R}$, we can think of the curve $y^2 = x^3 - n^2x$ heading off in an increasingly vertical direction as it approaches the line at infinity (see Fig. I.6). The points on the line at infinity correspond to the lines through the origin in the xy-plane, i.e., there is one for every possible slope y/x of such a line. As we move far out along our elliptic curve, we approach slope $y/x = \infty$, corresponding to the single point $(0, 1, 0)$ on the line at infinity. Notice that any elliptic curve $y^2 = f(x)$ similarly contains exactly one point at infinity $(0, 1, 0)$.

All of the usual concepts of calculus on curves $F(x, y) = 0$ in the xy-plane carry over to the corresponding projective curve $\tilde{F}(x, y, z) = 0$. Such notions as the tangent line at a point, points of inflection, smooth and singular points all depend only upon what is happening in a neighborhood of the point in question. And any point in $\mathbb{P}^2_{\mathbb{R}}$ has a large neighborhood which looks like an ordinary plane. More precisely, if we are interested in a point with nonzero z-coordinate, we can work in the usual xy-plane, where the curve has equation $F(x, y) = \tilde{F}(x, y, 1) = 0$. If we want to examine a point on the line $z = 0$, however, we put the triple in either the form $(x, 1, 0)$ or $(1, y, 0)$. In the former case, we think of it as a point on the curve $F(x, 1, z) = 0$

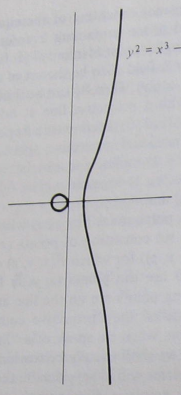

Figure I.6

in the xz-plane; and in the latter case as a point on the curve $F(1, y, z) = 0$ in the yz-plane.

For example, near the point at infinity $(0, 1, 0)$ on the elliptic curve $y^2z - x^3 + n^2xz^2$, all points have the form $(x, 1, z)$ with $z - x^3 + n^2xz^2 = 0$. The latter equation, in fact, gives us all points on the elliptic curve except for the three points $(0, 0, 1)$, $(\pm n, 0, 1)$ having zero y-coordinate (these are the three "points at infinity" if we think in terms of xz-coordinates).

Problems

1. Prove that if K is an infinite field and $F(x, y, z) \in K[x, y, z]$ satisfies $F(\lambda x, \lambda y, \lambda z) = \lambda^n F(x, y, z)$ for all $\lambda, x, y, z \in K$, then F is homogeneous, i.e., each monomial has total degree n. Give a counterexample if K is finite.

2. By a "line" in $\mathbb{P}^2_K$ we mean either the projective completion of a line in the xy-plane or the line at infinity. Show that a line in $\mathbb{P}^2_K$ has equation of the form $ax + by + cz = 0$, with $a, b, c \in K$ not all zero; and that two such equations determine the same line if and only if the two triples (a, b, c) differ by a multiple. Construct a 1-to-1 correspondence between lines in a copy of $\mathbb{P}^2_K$ with coordinates (x, y, z) and points in another copy of $\mathbb{P}^2_K$ with coordinates (a, b, c) and between points in the xyz-projective plane and lines in the abc-projective plane, such that a bunch of points are on the same line in the first projective plane if and only if the lines that correspond to them in the second projective plane all meet in the same point. The xyz-projective plane and the abc-projective plane are called the "duals" of each other.

3. How many points at infinity are on a parabola in $\mathbb{P}^2_{\mathbb{R}}$? an ellipse? a hyperbola?

4. Prove that any two nondegenerate conic sections in $\mathbb{P}^2_{\mathbb{R}}$ are equivalent to one another by some linear change of variables.

5. (a) If $\tilde{F}(x, y, z) \in K[x, y, z]$ is homogeneous of degree n, show that
$$x\frac{\partial \tilde{F}}{\partial x} + y\frac{\partial \tilde{F}}{\partial y} + z\frac{\partial \tilde{F}}{\partial z} = n\tilde{F}.$$
(b) If K has characteristic zero, show that a point $(x, y, z) \in \mathbb{P}^2_K$ is a non-smooth point on the curve C: $\tilde{F}(x, y, z) = 0$ if and only if the triple $(\partial\tilde{F}/\partial x, \partial\tilde{F}/\partial y, \partial\tilde{F}/\partial z)$ is $(0, 0, 0)$ at our particular (x, y, z). Give a counterexample if char $K \neq 0$. In what follows, suppose that char $K = 0$, e.g., $K = \mathbb{R}$.
(c) Show that the tangent line to C at a smooth point (x_0, y_0, z_0) has equation $ax + by + cz = 0$, where
$$a = \left.\frac{\partial \tilde{F}}{\partial x}\right|_{(x_0, y_0, z_0)}, \quad b = \left.\frac{\partial \tilde{F}}{\partial y}\right|_{(x_0, y_0, z_0)}, \quad c = \left.\frac{\partial \tilde{F}}{\partial z}\right|_{(x_0, y_0, z_0)}.$$
(d) Prove that the condition that (x, y, z) be a smooth point on C does not depend upon the choice of coordinates, i.e., it does not change if we shift to $x'y'z'$-coordinates, where $(x' \ y' \ z') = (x \ y \ z)A$ with A an invertible 3×3 matrix. For example, if more than one of the coordinates are nonzero, it makes no difference which we choose to regard as the "z-coordinate", i.e., whether we look at C in the xy-plane, the xz-plane, or the yz-plane.
(e) Prove that the condition that a given line l be tangent to C at a smooth point (x, y, z) does not depend upon the choice of coordinates.

6. (a) Let $P_1 = (x_1, y_1, z_1)$ and $P_2 = (x_2, y_2, z_2)$ be two distinct points in $\mathbb{P}^2_K$. Show that the line joining P_1 and P_2 can be given in parametrized form as $sP_1 + tP_2$, i.e., $\{(sx_1 + tx_2, sy_1 + ty_2, sz_1 + tz_2) | s, t \in K\}$. Check that this linear map takes $\mathbb{P}^1_K$ (with coordinates s, t) bijectively onto the line $\overline{P_1P_2}$ in $\mathbb{P}^2_K$. What part of the line do you get by taking $s = 1$ and letting t vary?
(b) Suppose that $K = \mathbb{R}$ or $\mathbb{C}$. If the curve $F(x, y) = 0$ in the xy-plane is smooth at $P_1 = (x_1, y_1)$ with nonvertical tangent line, then we can expand the implicit function $y = f(x)$ in a Taylor series about $x = x_1$. The linear term gives the tangent line. If we subtract off the linear term, we obtain $f(x) - y_1 - f'(x_1)(x - x_1) = a_m(x - x_1)^m + \cdots$, where $a_m \neq 0, m \geq 2$. m is called the "order of tangency". We say that (x_1, y_1) is a point of inflection if $m > 2$, i.e., $f''(x_1) = 0$. (In the case $K = \mathbb{R}$, note that we are not requiring a change in concavity with this definition, e.g., $y = x^4$ has a point of inflection at $x = 0$.) Let $P_1 = (x_1, y_1, z_1)$, $z_1 \neq 0$, and let $l = \overline{P_1P_2}$ be tangent to the curve $F(x, y) = \tilde{F}(x, y, 1)$ at the smooth point P_1. Let $P_2 = (x_2, y_2, z_2)$. Show that m is the lowest power of t that occurs in $\tilde{F}(x_1 + tx_2, y_1 + ty_2, z_1 + tz_2) \in K[t]$.
(c) Show that m does not change if we make a linear change of variables in $\mathbb{P}^2_K$. For example, suppose that y_1 and z_1 are both nonzero, and we use the xz-plane instead of the xy-plane in parts (a) and (b).

7. Show that the line at infinity (with equation $z = 0$) is tangent to the elliptic curve $y^2 = f(x)$ at $(0, 1, 0)$, and that the point $(0, 1, 0)$ is a point of inflection on the curve.

§4. Doubly periodic functions

Let L be a lattice in the complex plane, by which we mean the set of all integral linear combinations of two given complex numbers ω_1 and ω_2, where ω_1 and ω_2 do not lie on the same line through the origin. For example, if $\omega_1 = i$ and $\omega_2 = 1$, we get the lattice of Gaussian integers $\{mi + n | m, n \in \mathbb{Z}\}$. It will turn out that the example of the lattice of Gaussian integers is intimately related to the elliptic curves $y^2 = x^3 - n^2x$ that come from the congruent number problem.

The fundamental parallelogram for ω_1, ω_2 is defined as
$$\Pi = \{a\omega_1 + b\omega_2 | 0 \leq a \leq 1, 0 \leq b \leq 1\}.$$
Since ω_1, ω_2 form a basis for $\mathbb{C}$ over $\mathbb{R}$, any number $x \in \mathbb{C}$ can be written in the form $x = a\omega_1 + b\omega_2$ for some $a, b \in \mathbb{R}$. Then x can be written as the sum of an element in the lattice $L = \{m\omega_1 + n\omega_2\}$ and an element in Π, and in only one way unless a or b happens to be an integer, i.e., the element of Π happens to lie on the boundary $\partial\Pi$.

We shall always take ω_1, ω_2 in clockwise order; that is, we shall assume that ω_1/ω_2 has positive imaginary part.

Notice that the choice of ω_1, ω_2 giving the lattice L is not unique. For example, $\omega_1' = \omega_1 + \omega_2$ and ω_2 give the same lattice. More generally, we can obtain new bases ω_1', ω_2' for the lattice L by applying a matrix with integer entries and determinant 1 (see Problem 1 below).

For a given lattice L, a meromorphic function on $\mathbb{C}$ is said to be an *elliptic function* relative to L if $f(z + l) = f(z)$ for all $l \in L$. Notice that it suffices to check this property for $l = \omega_1$ and $l = \omega_2$. In other words, an elliptic function is periodic with two periods ω_1 and ω_2. Such a function is determined by its values on the fundamental parallelogram Π; and its values on opposite points of the boundary of Π are the same, i.e., $f(a\omega_1 + \omega_2) = f(a\omega_1)$, $f(\omega_1 + b\omega_2) = f(b\omega_2)$. Thus, we can think of an elliptic function $f(z)$ as a function on the set Π with opposite sides glued together. This set (more precisely, "complex manifold") is known as a "torus". It looks like a donut.

Doubly periodic functions on the complex numbers are directly analogous to singly periodic functions on the real numbers. A function $f(x)$ on $\mathbb{R}$ which satisfies $f(x + n\omega) = f(x)$ is determined by its values on the interval $[0, \omega]$. Its values at 0 and ω are the same, so it can be thought of as a function on the interval $[0, \omega]$ with the endpoints glued together. The "real manifold" obtained by gluing the endpoints is simply a circle (see Fig. I.7).

Returning now to elliptic functions for a lattice L, we let $\mathscr{E}_L$ denote the set of such functions. We immediately see that $\mathscr{E}_L$ is a subfield of the field of all meromorphic functions, i.e., the sum, difference, product, or quotient of two elliptic functions is elliptic. In addition, the subfield $\mathscr{E}_L$ is closed under differentiation. We now prove a sequence of propositions giving some very special properties which any elliptic function must have. The condition that a meromorphic function be doubly periodic turns out to be much more

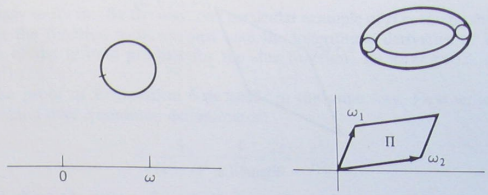

Figure I.7

restrictive than the analogous condition in the real case. The set of real-analytic periodic functions with given period is much "larger" than the set $\mathscr{E}_L$ of elliptic functions for a given period lattice L.

Proposition 3. *A function $f(z) \in \mathscr{E}_L$, $L = \{m\omega_1 + n\omega_2\}$, which has no pole in the fundamental parallelogram Π must be a constant.*

Proof. Since Π is compact, any such function must be bounded on Π, say by a constant M. But then $|f(z)| < M$ for all z, since the values of $f(z)$ are determined by the values on Π. By Liouville's theorem, a meromorphic function which is bounded on all of $\mathbb{C}$ must be a constant. □

Proposition 4. *With the same notation as above, let $\alpha + \Pi$ denote the translate of Π by the complex number α, i.e., $\{\alpha + z | z \in \Pi\}$. Suppose that $f(z) \in \mathscr{E}_L$ has no poles on the boundary C of $\alpha + \Pi$. Then the sum of the residues of $f(z)$ in $\alpha + \Pi$ is zero.*

Proof. By the residue theorem, this sum is equal to
$$\frac{1}{2\pi i}\int_C f(z)dz.$$
But the integral over opposite sides cancel, since the values of $f(z)$ at corresponding points are the same, while dz has opposite signs, because the path of integration is in opposite directions on opposite sides (see Fig. I.8). Thus, the integral is zero, and so the sum of residues is zero. □

Notice that, since a meromorphic function can only have finitely many poles in a bounded region, it is always possible to choose an α such that the boundary of $\alpha + \Pi$ misses the poles of $f(z)$. Also note that Proposition 4 immediately implies that a nonconstant $f(z) \in \mathscr{E}_L$ must have at least two poles (or a multiple pole), since if it had a single simple pole, then the sum of residues would not be zero.

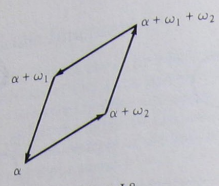

Figure I.8

Proposition 5. *Under the conditions of Proposition* 4, *suppose that* $f(z)$ *has no zeros or poles on the boundary of* $\alpha + \Pi$. *Let* $\{m_i\}$ *be the orders of the various zeros in* $\alpha + \Pi$, *and let* $\{n_j\}$ *be the orders of the various poles. Then* $\Sigma m_i = \Sigma n_j$.

PROOF. Apply Proposition 4 to the elliptic function $f'(z)/f(z)$. Recall that the logarithmic derivative $f'(z)/f(z)$ has a pole precisely where $f(z)$ has a zero or pole, such a pole is simple, and the residue there is equal to the order of zero or pole of the original $f(z)$ (negative if a pole). (Recall the argument: If $f(z) = c_m(z-a)^m + \cdots$, then $f'(z) = c_m m(z-a)^{m-1} + \cdots$, and so $f'(z)/f(z) = m(z-a)^{-1} + \cdots$.) Thus, the sum of the residues of $f'(z)/f(z)$ is $\Sigma m_i - \Sigma n_j = 0$. □

We now define what will turn out to be a key example of an elliptic function relative to the lattice $L = \{m\omega_1 + n\omega_2\}$. This function is called the Weierstrass $\wp$-function. It is denoted $\wp(z; L)$ or $\wp(z; \omega_1, \omega_2)$, or simply $\wp(z)$ if the lattice is fixed throughout the discussion. We set

$$\wp(z) = \wp(z; L) \underset{\text{def}}{=} \frac{1}{z^2} + \sum_{\substack{l \in L \\ l \neq 0}} \left(\frac{1}{(z-l)^2} - \frac{1}{l^2} \right). \tag{4.1}$$

Proposition 6. *The sum in* (4.1) *converges absolutely and uniformly for* z *in any compact subset of* $\mathbb{C} - L$.

PROOF. The sum in question is taken over a two-dimensional lattice. The proof of convergence will be rather routine if we keep in mind a one-dimensional analog. If instead of L we take the integers $\mathbb{Z}$, and instead of reciprocal squares we take reciprocals, we obtain a real function $f(x) = \frac{1}{x} + \Sigma \frac{1}{x-l} + \frac{1}{l}$, where the sum is over nonzero $l \in \mathbb{Z}$. To prove absolute and uniform convergence in any compact subset of $\mathbb{R} - \mathbb{Z}$, first write the summand as $x/(l(x-l))$, and then use a comparison test, showing that the series in question basically has the same behavior as l^{-2}. More precisely, use the following lemma: if Σb_l is a convergent sum of positive terms (all our sums being over nonzero $l \in \mathbb{Z}$), and if $\Sigma f_l(x)$ has the property that $|f_l(x)/b_l|$ approaches a finite limit as $l \to \pm\infty$, uniformly for x in some set, then the sum $\Sigma f_l(x)$ converges absolutely and uniformly for x in that set. The details

are easy to fill in. (By the way, our particular example of $f(x)$ can be shown to be the function $\pi \cot \pi x$; just take the logarithmic derivative of both sides of the infinite product for the sine function: $\sin \pi x = \pi x \Pi_{n=1}^{\infty}(1 - x^2/n^2)$.)

The proof of Proposition 6 proceeds in the same way. First write the summand over a common denominator:

$$\frac{1}{(z-l)^2} - \frac{1}{l^2} = \frac{2z - z^2/l}{(z-l)^2 l}.$$

Then show absolute and uniform convergence by comparison with $|l|^{-3}$, where the sum is taken over all nonzero $l \in L$. More precisely, Proposition 6 will follow from the following two lemmas.

Lemma 1. *If* Σb_l *is a convergent sum of positive terms, where the sum is taken over all nonzero elements in the lattice* L, *and if* $\Sigma f_l(z)$ *has the property that* $|f_l(z)/b_l|$ *approaches a finite limit as* $|l| \to \infty$, *uniformly for* z *in some subset of* $\mathbb{C}$, *then the sum* $\Sigma f_l(z)$ *converges absolutely and uniformly for* z *in that set.*

Lemma 2. $\Sigma |l|^{-s}$ *converges if* $s > 2$.

The proof of Lemma 1 is routine, and will be omitted. We give a sketch of the proof of Lemma 2. We split the sum into sums over l satisfying $n - 1 < |l| \leq n$, as $n = 1, 2, \ldots$. It is not hard to show that the number of l in that annulus has order of magnitude n. Thus, the sum in the lemma is bounded by a constant times $\Sigma_{n=1}^{\infty} n \cdot n^{-s} = \Sigma n^{1-s}$, and the latter sum converges for $s - 1 > 1$.

This concludes the proof of Proposition 6. □

Proposition 7. $\wp(z) \in \mathscr{E}_L$, *and its only pole is a double pole at each lattice point.*

PROOF. The same argument as in the proof of Proposition 6 shows that for any fixed $l \in L$, the function $\wp(z) - (z-l)^{-2}$ is continuous at $z = l$. Thus, $\wp(z)$ is a meromorphic function with a double pole at all lattice points and no other poles. Next, note that $\wp(z) = \wp(-z)$, because the right side of (4.1) remains unchanged if z is replaced by $-z$ and l is replaced by $-l$; but summing over $l \in L$ is the same as summing over $-l \in L$.

To prove double periodicity, we look at the derivative. Differentiating (4.1) term-by-term, we obtain:

$$\wp'(z) = -2 \sum_{l \in L} \frac{1}{(z-l)^3}.$$

Now $\wp'(z)$ is obviously doubly periodic, since replacing z by $z + l_0$ for some fixed $l_0 \in L$ merely rearranges the terms in the sum. Thus, $\wp'(z) \in \mathscr{E}_L$. To prove that $\wp(z) \in \mathscr{E}_L$, it suffices to show that $\wp(z + \omega_i) - \wp(z) = 0$ for $i = 1, 2$. We prove this for $i = 1$; the identical argument applies to $i = 2$.

Since the derivative of the function $\wp(z + \omega_1) - \wp(z)$ is $\wp'(z + \omega_1) - \wp'(z) = 0$, we must have $\wp(z + \omega_1) - \wp(z) = C$ for some constant C. But substituting $z = -\frac{1}{2}\omega_1$ and using the fact that $\wp(z)$ is an even function, we conclude that $C = \wp(\frac{1}{2}\omega_1) - \wp(-\frac{1}{2}\omega_1) = 0$. This concludes the proof. □

Notice that the double periodicity of $\wp(z)$ was not immediately obvious from the definition (4.1).

Since $\wp(z)$ has exactly one double pole in a fundamental domain of the form $\alpha + \Pi$, by Proposition 5 it has exactly two zeros there (or one double zero). The same is true of any elliptic function of the form $\wp(z) - u$, where u is a constant. It is not hard to show (see the problems below) that $\wp(z)$ takes every value $u \in \mathbb{C} \cup \{\infty\}$ exactly twice on the torus (i.e., a fundamental parallelogram with opposite sides glued together), counting multiplicity (which means the order of zero of $\wp(z) - u$); and that the values assumed with multiplicity two are ∞, $e_1 \underset{\text{def}}{=} \wp(\omega_1/2)$, $e_2 \underset{\text{def}}{=} \wp(\omega_2/2)$, and $e_3 \underset{\text{def}}{=} \wp((\omega_1 + \omega_2)/2)$. Namely, $\wp(z)$ has a double pole at 0, while the other three points are the zeros of $\wp'(z)$.

§5. The field of elliptic functions

Proposition 7 gives us a concrete example of an elliptic function. Just as $\sin x$ and $\cos x$ play a basic role in the theory of periodic functions on $\mathbb{R}$, because of Fourier expansion, similarly the functions $\wp(z)$ and $\wp'(z)$ play a fundamental role in the study of elliptic functions. But unlike in the real case, we do not even need infinite series to express an arbitrary elliptic function in terms of these two basic ones.

Proposition 8. $\mathscr{E}_L = \mathbb{C}(\wp, \wp')$, *i.e., any elliptic function for* L *is a rational expression in* $\wp(z; L)$ *and* $\wp'(z; L)$. *More precisely, given* $f(z) \in \mathscr{E}_L$, *there exist two rational functions* $g(X)$, $h(X)$ *such that* $f(z) = g(\wp(z)) + \wp'(z)h(\wp(z))$.

PROOF. If $f(z)$ is an elliptic function for L, then so are the two even functions

$$\frac{f(z) + f(-z)}{2} \quad \text{and} \quad \frac{f(z) - f(-z)}{2\wp'(z)}.$$

Since $f(z)$ is equal to the first of these functions plus $\wp'(z)$ times the second, to prove Proposition 8 it suffices to prove

Proposition 9. *The subfield* $\mathscr{E}_L^+ \subset \mathscr{E}_L$ *of even elliptic functions for* L *is generated by* $\wp(z)$, *i.e.,* $\mathscr{E}_L^+ = \mathbb{C}(\wp)$.

PROOF. The idea of the proof is to cook up a function which has the same zeros and poles as $f(z)$ using only functions of the form $\wp(z) - u$ with u a constant.

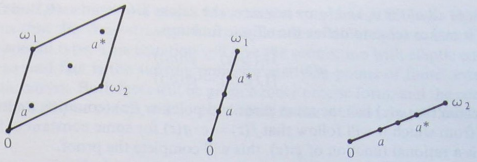

Figure I.9

The ratio of $f(z)$ to such a function is an elliptic function with no poles, and so must be a constant, by Proposition 3.

Let $f(z) \in \mathscr{E}_L^+$. We first list the zeros and poles of $f(z)$. But we must do this carefully, in a special way. Let Π' be a fundamental parallelogram with two sides removed: $\Pi' = \{a\omega_1 + b\omega_2 | 0 \leq a < 1, 0 \leq b < 1\}$. Then every point in $\mathbb{C}$ differs by a lattice element from *exactly* one point in Π'; that is, Π' is a set of coset representatives for the additive group of complex numbers modulo the subgroup L. We will list zeros and poles in Π', omitting 0 from our list (even if it happens to be a zero or pole of $f(z)$). Each zero or pole will be listed as many times as its multiplicity. However, only "half" will be listed; that is, they will be arranged in pairs, with only one taken from each pair. We now give the details. We describe the method of listing zeros; the method of listing poles is exactly analogous.

First suppose that $a \in \Pi'$, $a \neq 0$, is a zero of $f(z)$ which is *not* half of a lattice point, i.e., $a \neq \omega_1/2$, $\omega_2/2$, or $(\omega_1 + \omega_2)/2$. Let $a^* \in \Pi'$ be the point "symmetric" to a, i.e., $a^* = \omega_1 + \omega_2 - a$ if a is in the interior of Π', while $a^* = \omega_1 - a$ or $a^* = \omega_2 - a$ if a is on one of the two sides (see Fig. I.9). If a is a zero of order m, we claim that the symmetric point a^* is also a zero of order m. This follows from the double periodicity and the evenness of $f(z)$. Namely, we have $f(a^* - z) = f(-a - z)$ by double periodicity, and this is equal to $f(a + z)$ because $f(z)$ is an even function. Thus, if $f(a + z) = a_m z^m +$ higher terms, it follows that $f(a^* + z) = a_m(-z)^m +$ higher terms, i.e., a^* is a zero of order m.

Now suppose that $a \in \Pi'$ is a zero of $f(z)$ which *is* half of a lattice point; for example, suppose that $a = \omega_1/2$. In this case we claim that the order of zero m is even. If $f(a + z) = f(\frac{1}{2}\omega_1 + z) = a_m z^m +$ higher terms, then $f(\frac{1}{2}\omega_1 - z) = f(-\frac{1}{2}\omega_1 + z) = f(\frac{1}{2}\omega_1 + z)$ by double periodicity and evenness. Thus, $a_m z^m +$ higher terms $= a_m(-z)^m +$ higher terms, and so m is even.

We are now ready to list the zeros and poles of $f(z)$. Let $\{a_i\}$ be a list of the zeros of $f(z)$ in Π' which are not half-lattice points, each taken as many times as the multiplicity of zero there, but only one taken from each pair of symmetrical zeros a, a^*; in addition, if one of the three nonzero half-lattice points in Π' is a zero of $f(z)$, include it in the list half as many times as its multiplicity. Let $\{b_j\}$ be a list of the nonzero poles of $f(z)$ in Π', counted in the same way as the zeros (i.e., "only half" of them appear).

Since all of the a_i and b_j are nonzero, the values $\wp(a_i)$ and $\wp(b_j)$ are finite, and it makes sense to define the elliptic function

$$g(z) = \frac{\Pi_i(\wp(z) - \wp(a_i))}{\Pi_j(\wp(z) - \wp(b_j))}.$$

We claim that $g(z)$ has the same zeros and poles as $f(z)$ (counting multiplicity), from which it will follow that $f(z) = c \cdot g(z)$ for some constant c. Since $g(z)$ is a rational function of $\wp(z)$, this will complete the proof.

To prove this claim, we first examine nonzero points in Π'. Since 0 is the only pole in the numerator or denominator of $g(z)$, it follows that the nonzero zeros of $g(z)$ must come from the zeros of $\wp(z) - \wp(a_i)$, while the nonzero poles of $g(z)$ must come from the zeros of $\wp(z) - \wp(b_j)$. But we know (see problems below) that $\wp(z) - u$ (for constant u) has a double zero at $z = u$ if u is a half-lattice point, and otherwise has a pair of simple zeros at u and the symmetric point u^*. These are the only zeros of $\wp(z) - u$ in Π'. By our construction of the a_i and b_j, we see that $g(z)$ and $f(z)$ have the same order of zero or pole everywhere in Π', with the possible exception of the point 0. So it merely remains to show that they have the same order of zero or pole at 0. But this will follow automatically by Proposition 5. Namely, choose α so that no lattice point and no zero or pole of $f(z)$ or $g(z)$ is on the boundary of $\alpha + \Pi$. Then $\alpha + \Pi$ will contain precisely one lattice point l. We know that $f(z)$ and $g(z)$ have the same orders of zeros and poles everywhere in $\alpha + \Pi$ with the possible exception of l. Let m_f denote the order of zero of $f(z)$ at l (m_f is negative if there is a pole), and let m_g denote the analogous order for $g(z)$. Then

$$m_f + (\text{total of orders of zeros of } f) - (\text{total of orders of poles of } f)$$
$$= m_g + (\text{total of orders of zeros of } g) - (\text{total of orders of poles of } g).$$

Since the corresponding terms in parentheses on both sides of the equality are equal, we conclude that $m_f = m_g$. Thus, Proposition 5 tells us that when we know that two elliptic functions have the same order of zero or pole everywhere but possibly at one point in the fundamental parallelogram, then that one point is carried along automatically. This concludes the proof of Proposition 9. □

The proof of Propositions 8 and 9 was constructive, i.e., it gives us a prescription for expressing a given elliptic function in terms of $\wp(z)$ once we know its zeros and poles. Without doing any more work, for example, we can immediately conclude that:

(1) the even elliptic function $\wp'(z)^2$ is a cubic polynomial in $\wp(z)$ (because $\wp'(z)$ has a triple pole at 0 and three simple zeros, hence there are three a_i's and no b_j's);
(2) the even elliptic function $\wp(Nz)$ (for any fixed positive integer N) is a rational function in $\wp(z)$.

Both of these facts will play a fundamental role in what follows. The first tells us that the Weierstrass $\wp$-function satisfies a differential equation of a very special type. This equation will give the connection with elliptic curves. The second fact is the starting point for studying points of finite order on elliptic curves. Both facts will be given a more precise form, and the connection with elliptic curves will be developed, in the sections that follow.

Problems

1. Prove that the lattice $L = \{m\omega_1 + n\omega_2\}$ and the lattice $L' = \{m\omega_1' + n\omega_2'\}$ are the same if and only if there is a 2×2 matrix A with integer entries and determinant ± 1 such that $\omega' = A\omega$ (where ω denotes the column vector with entries ω_1, ω_2). If the pairs ω_1, ω_2 and ω_1', ω_2' are each listed in clockwise order, show that $\det A = +1$.

2. Let $\mathbb{C}/L$ denote the quotient of the additive group of complex numbers by the subgroup $L = \{m\omega_1 + n\omega_2\}$. Then $\mathbb{C}/L$ is in one-to-one correspondence with the fundamental parallelogram Π with opposite sides glued together.
   (a) Let C be the circle group (the unit circle in the complex plane). Give a continuous group isomorphism from $\mathbb{C}/L$ to the product of C with itself.
   (b) How many points of order N or a divisor of N are there in the group $\mathbb{C}/L$?
   (c) Show that the set of subgroups of prime order p in $\mathbb{C}/L$ is in one-to-one correspondence with the points of $\mathbb{P}^1_{\mathbb{F}_p}$ (where $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$). How many are there?

3. Let $s = 2, 3, 4, \ldots$. Fix a positive integer N, and let $f\colon \mathbb{Z} \times \mathbb{Z} \to \mathbb{C}$ be any function of period N, i.e., $f(m + N, n) = f(m, n)$ and $f(m, n + N) = f(m, n)$. Suppose that $f(0, 0) = 0$. If $s = 2$, further suppose that $\Sigma f(m, n) = 0$, where the sum is over $0 \le m, n < N$. Define a function

$$F_s(\omega_1, \omega_2) = \sum_{m,n \in \mathbb{Z}} \frac{f(m, n)}{(m\omega_1 + n\omega_2)^s}.$$

   (a) Prove that this sum converges absolutely if $s > 2$ and conditionally if $s = 2$ (in the latter case, take the sum over m and n in nondecreasing order of $|m\omega_1 + n\omega_2|$).
   (b) Express $F_s(\omega_1, \omega_2)$ in terms of the values of $\wp(z; \omega_1, \omega_2)$ or a suitable derivative evaluated at values of $z \in \Pi$ for which $Nz \in L$ (see Problem 2(b)).

4. Show that for any fixed u, the elliptic function $\wp(z) - u$ has exactly two zeros (or a single double zero). Use the fact that $\wp'(z)$ is odd to show that the zeros of $\wp'(z)$ are precisely $\omega_1/2$, $\omega_2/2$, and $(\omega_1 + \omega_2)/2$, and that the values $e_1 = \wp(\omega_1/2)$, $e_2 = \wp(\omega_2/2)$, $e_3 = \wp((\omega_1 + \omega_2)/2)$ are the values of u for which $\wp(z) - u$ has a double zero. Why do you know that e_1, e_2, e_3 are distinct? Thus, the Weierstrass $\wp$-function gives a two-to-one map from the torus (the fundamental parallelogram Π with opposite sides glued together) to the Riemann sphere $\mathbb{C} \cup \{\infty\}$ except over the four "branch points" e_1, e_2, e_3, ∞, each of which has a single preimage in $\mathbb{C}/L$.

5. Using the proof of Proposition 9, without doing any computations, what can you say about how the second derivative $\wp''(z)$ can be expressed in terms of $\wp(z)$?

§6. Elliptic curves in Weierstrass form

As remarked at the end of the last section, from the proof of Proposition 9 we can immediately conclude that the square of $\wp'(z)$ is equal to a cubic polynomial in $\wp(z)$. More precisely, we know that $\wp'(z)^2$ has a double zero at $\omega_1/2$, $\omega_2/2$, and $(\omega_1 + \omega_2)/2$ (see Problem 4 of §5). Hence, these three numbers are the a_i's, and we have

$$\wp'(z)^2 = C(\wp(z) - \wp(\omega_1/2))(\wp(z) - \wp(\omega_2/2))(\wp(z) - \wp((\omega_1 + \omega_2)/2))$$
$$= C(\wp(z) - e_1)(\wp(z) - e_2)(\wp(z) - e_3),$$

where C is some constant. It is easy to find C by comparing the coefficients of the lowest power of z in the Laurent expansion at the origin. Recall that $\wp(z) - z^{-2}$ is continuous at the origin, as is $\wp'(z) + 2z^{-3}$. Thus, the leading term on the left is $(-2z^{-3})^2 = 4z^{-6}$, while on the right it is $C(z^{-2})^3 = Cz^{-6}$. We conclude that $C = 4$. That is, $\wp(z)$ satisfies the differential equation

$$\wp'(z)^2 = f(\wp(z)), \qquad \text{where} \quad f(x) = 4(x - e_1)(x - e_2)(x - e_3) \in \mathbb{C}[x]. \tag{6.1}$$

Notice that the cubic polynomial f has distinct roots (see Problem 4 of §5).

We now give another independent derivation of the differential equation for $\wp(z)$ which uses only Proposition 3 from §4. Suppose that we can find a cubic polynomial $f(x) = ax^3 + bx^2 + cx + d$ such that the Laurent expansion at 0 of the elliptic function $f(\wp(z))$ agrees with the Laurent expansion of $\wp'(z)^2$ through the negative powers of z. Then the difference $\wp'(z)^2 - f(\wp(z))$ would be an elliptic function with no pole at zero, or in fact anywhere else (since $\wp(z)$ and $\wp'(z)$ have a pole only at zero). By Proposition 3, this difference is a constant; and if we suitably choose d, the constant term in $f(x)$, we can make this constant zero.

To carry out this plan, we must expand $\wp(z)$ and $\wp'(z)^2$ near the origin. Since both are even functions, only even powers of z will appear.

Let c be the minimum absolute value of nonzero lattice points l. We shall take $r < 1$, and assume that z is in the disc of radius rc about the origin. For each nonzero $l \in L$, we expand the term corresponding to l in the definition (4.1) of $\wp(z)$. We do this by differentiating the geometric series $1/(1 - x) = 1 + x + x^2 + \cdots$ and then substituting z/l for x:

$$\frac{1}{(1 - z/l)^2} = 1 + 2\frac{z}{l} + 3\frac{z^2}{l^2} + 4\frac{z^3}{l^3} + \cdots.$$

If we now subtract 1 from both sides, divide both sides by l^2, and then substitute in (4.1), we obtain

$$\wp(z) = \frac{1}{z^2} + \sum_{\substack{l \in L \\ l \neq 0}} 2\frac{z}{l^3} + 3\frac{z^2}{l^4} + 4\frac{z^3}{l^5} + \cdots + (k - 1)\frac{z^{k-2}}{l^k} + \cdots.$$

We claim that this double series is absolutely convergent for $|z| < rc$, in which case the following reversal of the order of summation will be justified:

$$\wp(z) = \frac{1}{z^2} + 3G_4 z^2 + 5G_6 z^4 + 7G_8 z^6 + \cdots, \tag{6.2}$$

where for $k > 2$ we denote

$$G_k = G_k(L) = G_k(\omega_1, \omega_2) \underset{\text{def}}{=} \sum_{\substack{l \in L \\ l \neq 0}} l^{-k} = \sum_{\substack{m,n \in \mathbb{Z} \\ \text{not both } 0}} \frac{1}{(m\omega_1 + n\omega_2)^k} \tag{6.3}$$

(notice that the G_k are zero for odd k, since the term for l cancels the term for $-l$; as we expect, only even powers of z occur in the expansion (6.2)). To check the claim of absolute convergence of the double series, we write the sum of the absolute values of the terms in the inner sum in the form (recall: $|z| < r|l|$):

$$2|z| \cdot |l|^{-3} \cdot \left(1 + \frac{3}{2}r + \frac{4}{2}r^2 + \frac{5}{2}r^3 + \cdots\right) < \frac{2|z|}{(1 - r)^2}\frac{1}{|l|^3},$$

and then use Lemma 2 from the proof of Proposition 6.

We now use (6.2) to compute the first few terms in the expansions of $\wp(z)$, $\wp(z)^2$, $\wp(z)^3$, $\wp'(z)$, and $\wp'(z)^2$, as follows:

$$\wp'(z) = -\frac{2}{z^3} + 6G_4 z + 20G_6 z^3 + 42G_8 z^5 + \cdots; \tag{6.4}$$

$$\wp'(z)^2 = \frac{4}{z^6} - 24G_4\frac{1}{z^2} - 80G_6 + (36G_4^2 - 168G_8)z^2 + \cdots; \tag{6.5}$$

$$\wp(z)^2 = \frac{1}{z^4} + 6G_4 + 10G_6 z^2 + \cdots; \tag{6.6}$$

$$\wp(z)^3 = \frac{1}{z^6} + 9G_4\frac{1}{z^2} + 15G_6 + (21G_8 + 27G_4^2)z^2 + \cdots. \tag{6.7}$$

Recall that we are interested in finding coefficients a, b, c, d of a cubic $f(x) = ax^3 + bx^2 + cx + d$ such that

$$\wp'(z)^2 = a\wp(z)^3 + b\wp(z)^2 + c\wp(z) + d,$$

and we saw that it suffices to show that both sides agree in their expansion through the constant term. If we multiply equation (6.7) by a, equation (6.6) by b, equation (6.2) by c, and then add them all to the constant d, and finally equate the coefficients of z^{-6}, z^{-4}, z^{-2} and the constant term to the corresponding coefficients in (6.5), we obtain successively:

$$a = 4; \qquad b = 0; \qquad -24G_4 = 4(9G_4) + c; \qquad -80G_6 = 4(15G_6) + d.$$

Thus, $c = -60G_4$, $d = -140G_6$. It is traditional to denote

$$g_2 = g_2(L) \underset{\text{def}}{=} 60G_4 = 60 \sum_{\substack{l \in L \\ l \neq 0}} l^{-4};$$
$$g_3 = g_3(L) \underset{\text{def}}{=} 140G_6 = 140 \sum_{\substack{l \in L \\ l \neq 0}} l^{-6}. \tag{6.8}$$

We have thereby derived a second form for the differential equation (6.1):

$$\wp'(z)^2 = f(\wp(z)), \qquad \text{where} \quad f(x) = 4x^3 - g_2 x - g_3 \in \mathbb{C}[x]. \tag{6.9}$$

Notice that if we were to continue comparing coefficients of higher powers of z in the expansion of both sides of (6.9), we would obtain relations between the various G_k (see Problems 4–5 below).

The differential equation (6.9) has an elegant and basic geometric interpretation. Suppose that we take the function from the torus $\mathbb{C}/L$ (i.e., the fundamental parallelogram Π with opposite sides glued) to $\mathbb{P}^2_{\mathbb{C}}$ defined by

$$z \mapsto (\wp(z), \wp'(z), 1) \quad \text{for} \quad z \neq 0;$$
$$0 \mapsto (0, 1, 0). \tag{6.10}$$

The image of any nonzero point z of $\mathbb{C}/L$ is a point in the xy-plane (with complex coordinates) whose x- and y-coordinates satisfy the relationship $y^2 = f(x)$ because of (6.9). Here $f(x) \in \mathbb{C}[x]$ is a cubic polynomial with distinct roots. Thus, every point z in $\mathbb{C}/L$ maps to a point on the elliptic curve $y^2 = f(x)$ in $\mathbb{P}^2_{\mathbb{C}}$. It is not hard to see that this map is a one-to-one correspondence between $\mathbb{C}/L$ and the elliptic curve (including its point at infinity). Namely, every x-value except for the roots of $f(x)$ (and infinity) has precisely two z's such that $\wp(z) = x$ (see Problem 4 of §5). The y-coordinates $y = \wp'(z)$ coming from these two z's are the two square roots of $f(x) = f(\wp(z))$. If, however, x happens to be a root of $f(x)$, then there is only one z value such that $\wp(z) = x$, and the corresponding y-coordinate is $y = \wp'(z) = 0$, so that again we are getting the solutions to $y^2 = f(x)$ for our given x.

Moreover, the map from $\mathbb{C}/L$ to our elliptic curve in $\mathbb{P}^2_{\mathbb{C}}$ is analytic, meaning that near any point of $\mathbb{C}/L$ it can be given by a triple of analytic functions. Near non-lattice points of $\mathbb{C}$ the map is given by $z \mapsto (\wp(z), \wp'(z), 1)$; and near lattice points the map is given by $z \mapsto (\wp(z)/\wp'(z), 1, 1/\wp'(z))$, which is a triple of analytic functions near L.

We have proved the following proposition.

Proposition 10. *The map* (6.10) *is an analytic one-to-one correspondence between* $\mathbb{C}/L$ *and the elliptic curve* $y^2 = 4x^3 - g_2(L)x - g_3(L)$ *in* $\mathbb{P}^2_{\mathbb{C}}$.

One might be interested in how the inverse map from the elliptic curve to $\mathbb{C}/L$ can be constructed. This can be done by taking path integrals of $dx/y = (4x^3 - g_2 x - g_3)^{-1/2} dx$ from a fixed starting point to a variable endpoint. The resulting integral depends on the path, but only changes by

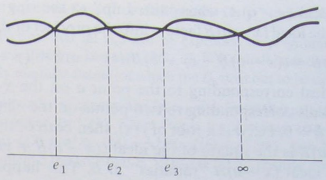

Figure I.10

a "period", i.e., a lattice element, if we change the path. We hence obtain a well-defined map to $\mathbb{C}/L$. See the exercises below for more details.

We conclude this section with a few words about an algebraic picture that is closely connected with the geometric setting of our elliptic curve. Recall from Proposition 8 that any elliptic function (meromorphic function on the torus $\mathbb{C}/L$) is a rational expression in $\wp(z)$ and $\wp'(z)$. Under our one-to-one correspondence in Proposition 10, such a function is carried over to a rational expression in x and y on the elliptic curve in the xy-plane (actually, in $\mathbb{P}^2_{\mathbb{C}}$). Thus, the field $\mathbb{C}(x, y)$ of rational functions on the xy-plane, when we restrict its elements to the elliptic curve $y^2 = f(x)$, and then "pull back" to the torus $\mathbb{C}/L$ by substituting $x = \wp(z)$, $y = \wp'(z)$, give us precisely the elliptic functions $\mathscr{E}_L$. Since the restriction of y^2 is the same as the restriction of $f(x)$, the field of functions obtained by restricting the rational functions in $\mathbb{C}(x, y)$ to the elliptic curve is the following quadratic extension of $\mathbb{C}(x)$: $\mathbb{C}(x)[y]/(y^2 - (4x^3 - g_2 x - g_3))$. Algebraically speaking, we form the quotient ring of $\mathbb{C}(x)[y]$ by the principal ideal corresponding to the equation $y^2 = f(x)$.

Geometrically, projection onto the x-coordinate gives us Fig. I.10. Two points on the elliptic curve map to one point on the projective line, except at four points (the point at infinity and the three points where $y = 0$), where the two "branches" are "pinched" together.

In algebraic geometry, one lets the field $F = \mathbb{C}(x)$ correspond to the complex line $\mathbb{P}^1_{\mathbb{C}}$, and the field $K = \mathbb{C}(x, y)/y^2 - (4x^3 - g_2 x - g_3)$ correspond to the elliptic curve in $\mathbb{P}^2_{\mathbb{C}}$. The rings $A = \mathbb{C}[x]$ and $B = \mathbb{C}[x, y]/y^2 - f(x)$ are the "rings of integers" in these fields. The maximal ideals in A are of the form $(x - a)A$; they are in one-to-one correspondence with $a \in \mathbb{C}$. A maximal ideal in B is of the form $(x - a)B + (y - b)B$ (where b is a square root of $f(a)$), and it corresponds to the point (a, b) on the elliptic curve.

$$\begin{array}{ccccc} K & \supset & B & \supset & (x-a)B + (y-b)B \qquad (b = \sqrt{f(a)}) \\ | & & | & & | \\ | & & | & & (x-a)B + (y+b)B \\ | & & | & & | \\ F & \supset & A & \supset & (x-a)A \end{array}$$

The maximal ideal $(x - a)A$, when "lifted up" to the ring B, is no longer prime. That is, the ideal $(x - a)B$ factors into the product of the two ideals:

$$(x - a)B = ((x - a)B + (y - b)B)((x - a)B + (y + b)B).$$

The maximal ideal corresponding to the point a on the x-line splits into two maximal ideals corresponding to two points on the elliptic curve. If it so happens that $b = 0$, i.e., a is a root of $f(x)$, then both of the ideals are the same, i.e., $(x - a)B$ is the square of the ideal $((x - a)B + yB)$. In that case we say that the ideal $(x - a)A$ "ramifies" in B. This happens at values a of the x-coordinate which come from only one point $(a, 0)$ on the elliptic curve. Thus, the above algebraic diagram of fields, rings and ideals is an exact mirror of the preceding geometric diagram.

We shall not go further than these *ad hoc* comments, since we shall not be using algebra geometric techniques in which follows. For a systematic introduction to algebraic geometry, see the textbooks by Shafarevich, Mumford, or Hartshorne.

PROBLEMS

1. (a) Let $L = \mathbb{Z}[i]$ be the lattice of Gaussian integers. Show that $g_3(L) = 0$ but that $g_2(L)$ is a nonzero real number.
   (b) Let $L = \mathbb{Z}[\omega]$, where $\omega = \frac{1}{2}(-1 + i\sqrt{3})$, be the lattice of integers in the quadratic imaginary field $\mathbb{Q}(\sqrt{-3})$. Show that $g_2(L) = 0$ but that $g_3(L)$ is a nonzero real number.
   (c) For any nonzero complex number c, let cL denote the lattice obtained by multiplying all lattice elements by c. Show that $g_2(cL) = c^{-4} g_2(L)$, and $g_3(cL) = c^{-6} g_3(L)$.
   (d) Prove that any elliptic curve $y^2 = 4x^3 - g_2 x - g_3$ with either g_2 or g_3 equal to zero, is of the form $y^2 = 4x^3 - g_2(L)x - g_3(L)$ for some lattice L. It can be shown that any elliptic curve is of that form for some lattice L. See, for example, [Whittaker & Watson 1958, §21.73]; also, we shall prove this much later as a corollary in our treatment of modular forms.

2. Recall that the discriminant of a polynomial $f(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_n = a_0(x - e_1)(x - e_2) \cdots (x - e_n)$ is $a_0^{n-1} \Pi_{i<j} (e_i - e_j)^2$. It is nonzero if and only if the roots are distinct. Since it is a symmetric homogeneous polynomial of degree $n(n-1)$ in the e_i's, it can be written as a polynomial in the elementary symmetric polynomials in the e_i's, which are $(-1)^i a_i / a_0$. Moreover, each monomial term $\Pi_i (a_i/a_0)^{m_i}$ has total "weight" $m_1 + 2m_2 + \cdots + nm_n$ equal to $n(n-1)$. Applying this to $f(x) = 4x^3 - g_2 x - g_3$, we see that the discriminant is equal to a polynomial in g_2, g_3 of weight six, i.e., it must be of the form $\alpha g_2^3 + \beta g_3^2$. Find α and β by computing $4^2 (e_1 - e_2)^2 (e_1 - e_3)^2 (e_2 - e_3)^2$ directly in the case $g_2 = 4$, $g_3 = 0$ and the case $g_2 = 0$, $g_3 = 4$.

3. Since the even elliptic function $\wp''(z)$ has a quadruple pole at zero and no other pole, you know in advance that it is equal to a quadratic polynomial in $\wp(z)$. Find this polynomial in two ways: (a) comparing coefficients of powers of z; (b) differentiating $\wp'^2 = 4\wp^3 - g_2 \wp - g_3$. Check that your answers agree.

4. Use either the equation for $\wp'^2$ or the equation for $\wp''$ to prove that $G_8 = \frac{3}{7} G_4^2$.

5. Prove by induction that all G_k's can be expressed as polynomials in G_4 and G_6 with rational coefficients, i.e., $G_k \in \mathbb{Q}[G_4, G_6]$. We shall later derive this fact again when we study modular forms (of which the G_k turn out to be examples).

6. Let $\omega_1 =$ it be purely imaginary, and let $\omega_2 = \pi$. Show that as t approaches infinity, $G_k(it, \pi)$ approaches $2\pi^{-k}\zeta(k)$, where $\zeta(s)$ is the Riemann zeta-function. Suppose we know that $\zeta(2) = \pi^2/6$, $\zeta(4) = \pi^4/90$, $\zeta(6) = \pi^6/945$. Use Problem 4 to find $\zeta(8)$. Use Problem 5 to show that $\pi^{-k}\zeta(k) \in \mathbb{Q}$ for all positive even integers k.

7. Find the limit of g_2 and g_3 for the lattice $L = \{mit + n\pi\}$ as $t \to \infty$.

8. Show that $v = \csc^2 z$ satisfies the differential equation $v'^2 = 4v^2(v - 1)$, and that the function

$$v = \csc^2 z - \tfrac{1}{3}$$

satisfies the differential equation $v'^2 = 4v^3 - \frac{4}{3}v - \frac{8}{27}$. What is the discriminant of the polynomial on the right? Now start with the infinite product formula for $\sin(\pi z)$, replace z by z/π, and take the logarithmic derivative and then the derivative once again to obtain an infinite sum for $\csc^2 z$. Then prove that

$$\lim_{t \to \infty} \wp(z; it, \pi) = \csc^2 z - \tfrac{1}{3}.$$

9. The purpose of this problem is to review the function $z = \log v$ for v complex, in the process providing a "dry run" for the problems that follow.
   (a) For v in a simply connected region of the complex plane that does not include the origin, define a function z of v by:

$$z = \int_1^v \frac{dt}{t},$$

   where the path from 1 to v is chosen arbitrarily, except that the same choice is made for all points in the region. (In other words, fix any path from 1 to v_0, and then to go to other v's use a path from v_0 to v that stays in the region.) Call this function $z = \log v$. Show that if a different path is chosen, the function changes by a constant value in the "lattice" $L = \{2\pi i m\}$; and that any lattice element can be added to the function by a suitable change of path. (L is actually only a lattice in the imaginary axis $\mathbb{R}i$, not a lattice in $\mathbb{C}$.)
   (b) Express dz/dv and dv/dz in terms of v.
   (c) If the function $v = e^z$ is defined by the usual series, use part (b) to show that e^z is the inverse function of $z = \log v$.
   (d) Show that the map e^z gives a one-to-one correspondence between $\mathbb{C}/L$ and $\mathbb{C} - \{0\}$. Under this one-to-one correspondence, the additive group law in $\mathbb{C}/L$ becomes what group law in $\mathbb{C} - \{0\}$?

10. Let L be a fixed lattice, set $g_2 = g_2(L)$, $g_3 = g_3(L)$, $\wp(z) = \wp(z; L)$. Let $u = f(z)$ be a function on a connected open region $R \subset \mathbb{C}$ which satisfies the differential equation $u'^2 = 4u^3 - g_2 u - g_3$. Prove that $u = \wp(z + \alpha)$ for some constant α.

11. Let $L = \{m\omega_1 + n\omega_2\}$ be a fixed lattice, and set $g_2 = g_2(L)$, $g_3 = g_3(L)$, $\wp(z) = \wp(z; L)$. Let R_1 be an unbounded simply connected open region in the complex plane which does not contain the roots e_1, e_2, e_3 of the cubic $4x^3 - g_2 x - g_3$.

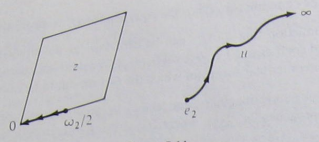

Figure I.11

For $u \in R_1$, define a function $z = g(u)$ by

$$z = g(u) = \int_u^\infty \frac{dt}{\sqrt{4t^3 - g_2 t - g_3}},$$

where a fixed branch of the square root is chosen as t varies in R_1. Note that the integral converges and is independent of the path in R_1 from u to ∞, since R_1 is simply connected. The function $z = g(u)$ can be analytically continued by letting R_2 be a simply connected region in $\mathbb{C} - \{e_1, e_2, e_3\}$ which overlaps with R_1. If $u \in R_2$, then choose $u_1 \in R_1 \cap R_2$, and set $z = g(u) = g(u_1) + \int_u^{u_1} (4t^3 - g_2 t - g_3)^{-1/2} dt$. This definition clearly does not depend on our choice of $u_1 \in R_1 \cap R_2$ or our path from u to u_1 in R_2. Continuing in this way, we obtain an analytic function which is multivalued, because our sequence of regions $R_1, R_2, R_3, \ldots$ can wind around e_1, e_2, or e_3.

(a) Express $(dz/du)^2$ and $(du/dz)^2$ in terms of u.

(b) Show that $u = \wp(z)$. In particular, when we wind around e_1, e_2, or e_3 the value of z can only change by something in L. Thus, $z = g(u)$ is well defined as an element in $\mathbb{C}/L$ for $u \in \mathbb{C} - \{e_1, e_2, e_3\}$. The function $z = g(u)$ then extends by continuity to e_1, e_2, e_3.

(c) Let C_1 be the path in the complex u-plane from e_2 to ∞ that is traced by $u = \wp(z)$ as z goes from $\omega_2/2$ to 0 along the side of Π (see Fig. I.11). Show that $\int_{C_1} (4t^3 - g_2 t - g_3)^{-1/2} dt = -\omega_2/2$ for a suitable branch of the square root.

(d) Let C_2 be the path that goes from ∞ to e_2 along C_1, winds once around e_2, and then returns along C_1 to ∞. Take the same branch of the square root as in part (c), and show that $\int_{C_2} (4t^3 - g_2 t - g_3)^{-1/2} dt = \omega_2$.

(e) Describe how the function $z = g(u)$ can be made to give all preimages of u under $u = \wp(z)$.

12. (a) Prove that all of the roots e_1, e_2, e_3 of $4x^3 - g_2 x - g_3$ are real if and only if g_2 and g_3 are real and $\Delta = g_2^3 - 27 g_3^2 > 0$.

(b) Suppose that the conditions in part (a) are met, and we order the e_i so that $e_2 > e_3 > e_1$. Show that we can choose the periods of L to be given by

$$\frac{1}{2}\omega_1 = i\int_{-\infty}^{e_1} \frac{dt}{\sqrt{g_3 + g_2 t - 4t^3}} \quad \text{and} \quad \frac{1}{2}\omega_2 = \int_{e_2}^{\infty} \frac{dt}{\sqrt{4t^3 - g_2 t - g_3}},$$

where we take the positive branch of the square root, and integrate along the real axis.

(c) With these assumptions about the location of the e_i on the real axis, describe how to change the path of integration and the branch of the square root in Problem 11 so as to get the other values of z for which $u = \wp(z)$, namely $\pm z + m\omega_1 + n\omega_2$.

13. Suppose that $g_2 = 4n^2$, $g_3 = 0$. Take e_1, e_2, e_3 so that $e_2 > e_3 > e_1$. What are e_1, e_2, e_3 in this case? Show that $\omega_1 = i\omega_2$, i.e., the lattice L is the Gaussian integer lattice expanded by a factor of ω_2. Show that as z travels along the straight line from $\omega_1/2$ to $\omega_1/2 + \omega_2$ the point $(x, y) = (\wp(z), \wp'(z))$ moves around the real points of the elliptic curve $y^2 = 4(x^3 - n^2 x)$ between $-n$ and 0; and as z travels along the straight line from 0 to ω_2 the point $(x, y) = (\wp(z), \wp'(z))$ travels through all the real points of this elliptic curve which are to the right of $(n, 0)$. Think of the "open" appearance of the latter path to be an optical illusion: the two ends are really "tied together" at the point at infinity (0, 1, 0).

14. (a) Show that $\displaystyle\int_0^1 \frac{t^n dt}{\sqrt{t(1-t)}} = \frac{\pi}{n!}\cdot\frac{1}{2}\cdot\frac{3}{2}\cdot\frac{5}{2}\cdots\left(n - \frac{1}{2}\right)$ for $n = 0, 1, 2, \ldots$.

(b) Under the conditions of Problem 12, with $e_2 > e_3 > e_1$, set $\lambda = \dfrac{e_3 - e_1}{e_2 - e_1} \in (0, 1)$. Derive the formula:

$$\omega_2 = \frac{1}{\sqrt{e_2 - e_1}} \int_0^1 \frac{dt}{\sqrt{t(1-t)(1-\lambda t)}}.$$

(c) Derive the formula $\omega_2 = \pi(e_2 - e_1)^{-1/2} F(\lambda)$, where

$$F(\lambda) = \sum_{n=0}^{\infty} \left[\frac{1}{2}\cdot\frac{3}{2}\cdot\frac{5}{2}\cdots\left(n - \frac{1}{2}\right)\right]^2 \frac{\lambda^n}{n!^2}.$$

The function $F(\lambda)$ is called a "hypergeometric series".

(d) Show that the hypergeometric series in part (c) satisfies the differential equation: $\lambda(1-\lambda)F''(\lambda) + (1 - 2\lambda)F'(\lambda) - \frac{1}{4}F(\lambda) = 0$.

§7. The addition law

In the last section we showed how the Weierstrass $\wp$-function gives a correspondence between the points of $\mathbb{C}/L$ and the points on the elliptic curve $y^2 = f(x) = 4x^3 - g_2(L)x - g_3(L)$ in $\mathbb{P}^2_{\mathbb{C}}$. We have an obvious addition law for points in $\mathbb{C}/L$, obtained from ordinary addition of complex numbers by dividing by the additive subgroup L, i.e., ordinary addition "modulo L". This is the two-dimensional analog of "addition modulo one" in the group $\mathbb{R}/\mathbb{Z}$.

We can use the correspondence between $\mathbb{C}/L$ and the elliptic curve to carry over the addition law to the points on the elliptic curve. That is, to add two points $P_1(x_1, y_1)$ and $P_2 = (x_2, y_2)$, by definition what we do is go back to the z-plane, find z_1 and z_2 such that $P_1 = (\wp(z_1), \wp'(z_1))$ and $P_2 = (\wp(z_2), \wp'(z_2))$, and then set $P_1 + P_2 = (\wp(z_1 + z_2), \wp'(z_1 + z_2))$. This is just a case of the general principle: whenever we have a one-to-one correspondence between elements of a commutative group and elements of some

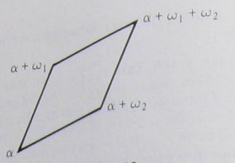

Figure I.12

other set, we can use this correspondence to define a commutative group law on that other set.

But the remarkable thing about the addition law we obtain in this way is that (1) there is a simple geometric interpretation of "adding" the points on the elliptic curve, and (2) the coordinates of $P_1 + P_2$ can be expressed directly in terms of x_1, x_2, y_1, y_2 by rather simple rational functions. The purpose of this section is to show how this is done.

We first prove a general lemma about elliptic functions.

Lemma. *Let $f(z) \in \mathscr{E}_L$. Let $\Pi = \{a\omega_1 + b\omega_2 | 0 \le a, b \le 1\}$ be a fundamental parallelogram for the lattice L, and choose α so that $f(z)$ has no zeros or poles on the boundary of $\alpha + \Pi$. Let $\{a_i\}$ be the zeros of $f(z)$ in $\alpha + \Pi$, each repeated as many times as its multiplicity, and let $\{b_j\}$ be the poles, each occurring as many times as its multiplicity. Then $\Sigma a_i - \Sigma b_j \in L$.*

PROOF. Recall that the function $f'(z)/f(z)$ has poles at the zeros and poles of $f(z)$, and its expansion near a zero a of order m is $m/(z-a) + \cdots$ (and near a pole b of order $-m$ the expansion is $-m/(z-b) + \cdots$). Then the function $zf'(z)/f(z)$ has the same poles, but, writing $z = a + (z - a)$, we see that the expansion starts out $am/(z-a)$. We conclude that $\Sigma a_i - \Sigma b_j$ is the sum of the residues of $zf'(z)/f(z)$ inside $\alpha + \Pi$. Let C be the boundary of $\alpha + \Pi$. By the residue theorem,

$$\sum a_i - \sum b_j = \frac{1}{2\pi i}\int_C \frac{zf'(z)}{f(z)} dz.$$

We first take the integral over the pair of opposite sides from α to $\alpha + \omega_2$ and from $\alpha + \omega_1$ to $\alpha + \omega_1 + \omega_2$ (see Fig. I.12). This part is equal to

$$\begin{aligned}
&\frac{1}{2\pi i}\left(\int_\alpha^{\alpha+\omega_2} z\frac{f'(z)}{f(z)}dz - \int_{\alpha+\omega_1}^{\alpha+\omega_1+\omega_2} z\frac{f'(z)}{f(z)}dz\right) \\
&= \frac{1}{2\pi i}\left(\int_\alpha^{\alpha+\omega_2} z\frac{f'(z)}{f(z)}dz - \int_\alpha^{\alpha+\omega_2} (z+\omega_1)\frac{f'(z)}{f(z)}dz\right) \\
&= -\omega_1 \frac{1}{2\pi i}\int_\alpha^{\alpha+\omega_2} \frac{f'(z)}{f(z)}dz.
\end{aligned}$$

Now make the change of variables $u = f(z)$, so that $f'(z)dz/f(z) = du/u$. Let C_1 be the closed path from $f(\alpha)$ to $f(\alpha + \omega_2) = f(\alpha)$ traced by $u = f(z)$ as z goes from α to $\alpha + \omega_2$. Then

$$\frac{1}{2\pi i}\int_\alpha^{\alpha+\omega_2} \frac{f'(z)}{f(z)}dz = \frac{1}{2\pi i}\int_{C_1} \frac{du}{u},$$

and this is some integer n, namely the number of times the closed path C_1 winds around the origin (counterclockwise). Thus, we obtain $-\omega_1 n$ for this part of our original integral. In the same way, we find that the integral over the remaining two sides of C is equal to $-\omega_2 m$ for some integer m. Thus, $\Sigma a_i - \Sigma b_j = -n\omega_1 - m\omega_2 \in L$, as desired. This proves the lemma. □

We are now ready to derive the geometrical procedure for adding two points on the elliptic curve $y^2 = f(x) = x^3 - g_2(L)x - g_3(L)$. For z in $\mathbb{C}/L$, let P_z be the corresponding point $P_z = (\wp(z), \wp'(z), 1)$, $P_0 = (0, 1, 0)$ on the elliptic curve. Suppose we want to add $P_{z_1} = (x_1, y_1)$ to $P_{z_2} = (x_2, y_2)$ to obtain the sum $P_{z_1+z_2} = (x_3, y_3)$. We would like to know how to go from the two points to their sum directly, without tracing the points back to the z-plane.

We first treat some special cases. The additive identity is, of course, the image of $z = 0$. Let 0 denote the point at infinity (0, 1, 0), i.e., the additive identity of our group of points. The addition is trivial if one of the points is 0, i.e., if z_1 or z_2 is zero. Next, suppose that P_{z_1} and P_{z_2} have the same x-coordinate but are not the same point. This means that $x_2 = x_1, y_2 = -y_1$. In this case $z_2 = -z_1$, because only "symmetric" values of z (values which are the negatives of each other modulo the lattice L) can have the same $\wp$-value. In this case, $P_{z_1} + P_{z_2} = P_0 = 0$, i.e., the two points are additive inverse to one another. Speaking geometrically, we say that two points of the curve which are on the same vertical line have sum 0. We further note that in the special situation of a point $P_{z_1} = P_{z_2}$ on the x-axis, we have $y_2 = -y_1 = 0$, and it is easy to check that we still have $P_{z_1} + P_{z_2} = 2P_{z_1} = 0$. We have proved:

Proposition 11. *The additive inverse of (x, y) is $(x, -y)$.*

Given two points $P_1 = P_{z_1} = (x_1, y_1)$ and $P_2 = P_{z_2} = (x_2, y_2)$ on the elliptic curve $y^2 = 4x^3 - g_2 x - g_3$ (neither the point at infinity 0), there is a line $l = \overline{P_1 P_2}$ joining them. If $P_1 = P_2$, we take l to be the tangent line to the elliptic curve at P_1. If l is a vertical line, then we saw that $P_1 + P_2 = 0$. Suppose that l is not a vertical line, and we want to find $P_1 + P_2 = P_3 = (x_3, y_3)$. Our basic claim is that $-P_3 = (x_3, -y_3)$ is the third point of intersection of the elliptic curve with l.

Write the equation of $l = \overline{P_1 P_2}$ in the form $y = mx + b$. A point (x, y) on l is on the elliptic curve if and only if $(mx + b)^2 = f(x) = 4x^3 - g_2 x - g_3$, that is, if and only if x is a root of the cubic $f(x) - (mx + b)^2$. This cubic

has three roots, each of which gives a point of intersection. If x is a double root or triple root, then l intersects the curve with multiplicity two or three at the point (x, y) (see Problem 6 of §I.3). In any case, the total number of points of intersection (counting multiplicity) is three.

Notice that vertical lines also intersect the curve in three points, including the point at infinity 0; and the line at infinity has a triple intersection at 0 (see Problem 7 of §I.3). Thus, any line in $\mathbb{P}^2_{\mathbb{C}}$ intersects the curve in three points. This is a special case of

Bezout's Theorem. *Let $\tilde{F}(x, y, z)$ and $\tilde{G}(x, y, z)$ be homogeneous polynomials of degree m and n, respectively, over an algebraically closed field K. Suppose that $\tilde{F}$ and $\tilde{G}$ have no common polynomial factor. Then the curves in $\mathbb{P}^2_K$ defined by $\tilde{F}$ and $\tilde{G}$ have mn points of intersection, counting multiplicities.*

For a more detailed discussion of multiplicity of intersection and a proof of Bezout's theorem, see, for example, Walker's book an algebraic curves [Walker 1978].

In our case $\tilde{F}(x, y, z) = y^2z - 4x^3 + g_2xz^2 + g_3z^3$ and $\tilde{G}(x, y, z) = y - mx - bz$.

Proposition 12. *If $P_1 + P_2 = P_3$, then $-P_3$ is the third point of intersection of $l = \overline{P_1P_2}$ with the elliptic curve. If $P_1 = P_2$, then by $\overline{P_1P_2}$ we mean the tangent line at P_1.*

PROOF. We have already treated the case when P_1 or P_2 is the point at infinity 0, and when $P_2 = -P_1$. So suppose that $l = \overline{P_1P_2}$ has the form $y = mx + b$. Let $P_1 = P_{z_1}$, $P_2 = P_{z_2}$. To say that a point $P_z = (\wp(z), \wp'(z))$ is on l means that $\wp'(z) = m\wp(z) + b$. The elliptic function $\wp'(z) - m\wp(z) - b$ has three poles and hence three zeros in $\mathbb{C}/L$. Both z_1 and z_2 are zeros. According to the lemma proved above, the sum of the three zeros and three poles is equal to zero modulo the lattice L. But the three poles are all at zero (where $\wp'(z)$ has a triple pole); thus, the third zero is $-(z_1 + z_2)$ modulo the lattice. Hence, the third point of intersection of l with the curve is $P_{-(z_1+z_2)} = -P_{z_3}$, as claimed.

The argument in the last paragraph is rigorous only if the three points of intersection of l with the elliptic curve are distinct, in which case a zero of $\wp'(z) - m\wp(z) - b$ corresponds exactly to a point of intersection P_z. Otherwise, we must show that a double or triple zero of the elliptic function always corresponds to a double or triple intersection, respectively, of l with the curve. That is, we must show that the two meanings of the term "multiplicity" agree: multiplicity of zero of the elliptic function of the variable z, and multiplicity of intersection in the xy-plane.

Let $z_1, z_2, -z_3$ be the three zeros of $\wp'(z) - m\wp(z) - b$, listed as many times as their multiplicity. Note that none of these three points is the negative of another one, since l is not a vertical line. Since $-z_1, -z_2, z_3$ are the three zeros of $\wp'(z) + m\wp(z) + b$, it follows that $\pm z_1, \pm z_2, \pm z_3$ are the six zeros of $\wp'(z)^2 - (m\wp(z) + b)^2 = f(\wp(z)) - (m\wp(z) + b)^2 = 4(\wp(z) - x_1)(\wp(z) - x_2)(\wp(z) - x_3)$, where x_1, x_2, x_3 are the roots of $f(x) - (mx + b)^2$. If, say, $\wp(z_1) = x_1$, then the multiplicity of x_1 depends upon the number of $\pm z_2, \pm z_3$ which equal $\pm z_1$. But this is precisely the number of $z_2, -z_3$ which equal z_1. Hence "multiplicity" has the same meaning in both cases.

This concludes the proof of Proposition 12. □

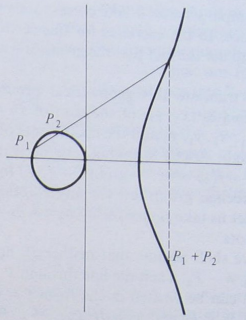

Figure I.13

Proposition 12 gives us Fig. I.13, which illustrates the group of real points on the elliptic curve $y^2 = x^3 - x$. To add two points P_1 and P_2, we draw the line joining them, find the third point of intersection of that line with the curve, and then take the symmetric point on the other side of the x-axis.

It would have been possible to define the group law in this geometrical manner in the first place, and prove directly that the axioms of an abelian group are satisfied. The hardest part would have been the associative law, which would have necessitated a deeper investigation of intersections of curves. In turns out that there is some flexibility in defining the group law. For example, any one of the eight points of inflection besides the point at infinity could equally well have been chosen as the identity. For details of this alternate approach, see [Walker 1978].

One disadvantage of our approach using $\wp(z)$ is that *a priori* it only applies to elliptic curves of the form $y^2 = 4x^3 - g_2(L)x - g_3(L)$ or curves that can be transformed to that form by a linear change of variables. (Note that the geometrical description of the group law will still give an abelian group law after a linear change of variables.) In actual fact, as was mentioned earlier and will be proved later, any elliptic curve over the complex numbers can be transformed to the Weierstrass form for some lattice L. We already know

that our favorite example $y^2 = x^3 - n^2x$ corresponds to a multiple of the Gaussian integer lattice. In the exercises for this section and the next, we shall allow ourselves to use the fact that the group law works for any elliptic curve.

It is not hard to translate this geometrical procedure into formulas expressing the coordinates (x_3, y_3) of the sum of $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ in terms of x_1, x_2, y_1, y_2 and the coefficients of the equation of the elliptic curve. Although, strictly speaking, our derivation was for elliptic curves in the form $y^2 = f(x) = 4x^3 - g_2(L)x - g_3(L)$ for some lattice L, the procedure gives an abelian group law for any elliptic curve $y^2 = f(x)$, as remarked above. So let us take $f(x) = ax^3 + bx^2 + cx + d \in \mathbb{C}[x]$ to be any cubic with distinct roots.

In what follows, we shall assume that neither P_1 nor P_2 is the point at infinity 0, and that $P_1 \neq -P_2$. Then the line through P_1 and P_2 (the tangent line at P_1 if $P_1 = P_2$) can be written in the form $y = mx + \beta$, where $m = (y_2 - y_1)/(x_2 - x_1)$ if $P_1 \neq P_2$ and $m = dy/dx|_{(x_1, y_1)}$ if $P_1 = P_2$. In the latter case we can express m in terms of x_1 and y_1 by implicitly differentiating $y^2 = f(x)$; we find that $m = f'(x_1)/2y_1$. In both cases the y-intercept is $\beta = y_1 - mx_1$.

Then x_3, the x-coordinate of the sum, is the third root of the cubic $f(x) - (mx + \beta)^2$, two of whose roots are x_1, x_2. Since the sum of the three roots is equal to minus the coefficient of x^2 divided by the leading coefficient, we have: $x_1 + x_2 + x_3 = -(b - m^2)/a$, and hence:

$$x_3 = -x_1 - x_2 - \frac{b}{a} + \frac{1}{a}\left(\frac{y_2 - y_1}{x_2 - x_1}\right)^2, \qquad \text{if} \quad P_1 \neq P_2; \tag{7.1}$$

$$x_3 = -2x_1 - \frac{b}{a} + \frac{1}{a}\left(\frac{f'(x_1)}{2y_1}\right)^2, \qquad \text{if} \quad P_1 = P_2. \tag{7.2}$$

The y-coordinate y_3 is the negative of the value $y = mx_3 + \beta$, i.e.,

$$y_3 = -y_1 + m(x_1 - x_3), \tag{7.3}$$

where x_3 is given by (7.1) and (7.2), and

$$\begin{aligned} m &= (y_2 - y_1)/(x_2 - x_1) \quad \text{if} \quad P_1 \neq P_2; \\ m &= f'(x_1)/2y_1 \qquad\qquad \text{if} \quad P_1 = P_2. \end{aligned} \tag{7.4}$$

If our elliptic curve is in Weierstrass form $y^2 = 4x^3 - g_2x - g_3$, then we have $a = 4$, $b = 0$, and $f'(x_1) = 12x_1^2 - g_2$ in the addition formulas (7.1)–(7.4).

In principle, we could have simply defined the group law by means of these formulas, and then verified algebraically that the axioms of a commutative group are satisfied. The hardest axiom to verify would be associativity. Tedious as this procedure would be, it would have one key advantage over

either the complex-analytic procedure (using $\wp(z)$) or the geometrical procedure. Namely, we would never have to use the fact that our field K over which the elliptic curve is defined is the complex numbers, or even that it has characteristic zero. That is, we would find that our formulas, which make sense over any field K of characteristic not equal to 2, give an abelian group law. That is, if $y^2 = f(x) = ax^3 + bx^2 + cx + d \in K[x]$ is the equation of an elliptic curve over K, and if we define $f'(x) = 3ax^2 + 2bx + c$, then any two points having coordinates in some extension of K can be added using the formulas (7.1)–(7.4). We shall make use of this fact in what follows, even though, strictly speaking, we have not gone through the tedious purely algebraic verification of the group laws.

PROBLEMS

1. Let $L \subset \mathbb{R}$ be the additive subgroup $\{m\omega\}$ of multiples of a fixed nonzero real number ω. Then the function $z \mapsto (\cos(2\pi z/\omega), \sin(2\pi z/\omega))$ gives a one-to-one analytic map of $\mathbb{R}/L$ onto the curve $x^2 + y^2 = 1$ in the real xy-plane. Show that ordinary addition in $\mathbb{R}/L$ carries over to a rational (actually polynomial) law for "adding" two points (x_1, y_1) and (x_2, y_2) on the unit circle; that is, the coordinates of the "sum" are polynomials in x_1, x_2, y_1, y_2. Thus, the rational addition law on an elliptic curve can be thought of as a generalization of the formulas for the sine and cosine of the sum of two angles.

2. (a) Simplify the expression for the x-coordinate of $2P$ in the case of the elliptic curve $y^2 = x^3 - n^2x$.
   (b) Let X, Y, Z be a rational right triangle with area n. Let P be the corresponding point on the curve $y^2 = x^3 - n^2x$ constructed in the text in §I.2. Let Q be the point constructed in Problem 2 of §I.2. Show that $P = 2Q$.
   (c) Prove that, if P is a point not of order 2 with rational coordinates on the curve $y^2 = x^3 - n^2x$, then the x-coordinate of $2P$ is the square of a rational number having even denominator. For example, the point $Q = ((41/7)^2, 720 \cdot 41/7^3)$ on the curve $y^2 = x^3 - 31^2x$ is not equal to twice a point P having rational coordinates. (In this problem, recall: n is always squarefree.)

3. Describe geometrically: (a) the four points of order two on an elliptic curve; (b) the nine points of order three; (c) how to find the twelve points of order four which are not of order two; (d) what the associative law of addition says about a certain configuration of lines joining points on the elliptic curve (draw a picture).

4. (a) How many points of inflection are there on an elliptic curve besides the point at infinity? Notice that they occur in symmetric pairs. Find an equation for their x-coordinates.
   (b) In the case of the elliptic curve $y^2 = x^3 - n^2x$ find an explicit formula for these x-coordinates. Show that they are never rational (for any n).

5. Given a point Q on an elliptic curve, how many points P are there such that $2P = Q$? Describe geometrically how to find them.

6. Show that if K is any subfield of $\mathbb{C}$ containing g_2 and g_3, then the points on the elliptic curve $y^2 = 4x^3 - g_2x - g_3$ whose coordinates are in K form a subgroup

of the group of all points. More generally, show that this is true for the elliptic curve $y^2 = f(x)$ if $f(x) \in K[x]$.

7. Consider the subgroup of all points on $y^2 = x^3 - n^2x$ with real coordinates. How many points in this subgroup are of order 2? 3? 4? Describe geometrically where these points are located.

8. Same as Problem 7 for the elliptic curve $y^2 = x^3 - a$, $a \in \mathbb{R}$.

9. If $y^2 = f(x)$ is an elliptic curve in which $f(x)$ has real coefficients, show that the group of points with real coordinates is isomorphic to (a) $\mathbb{R}/\mathbb{Z}$ if $f(x)$ has only one real root; (b) $\mathbb{R}/\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ if $f(x)$ has three real roots.

10. Letting a approach zero in Problem 8, show that for the curve $y^2 = x^3$ the same geometric procedure for finding $P_1 + P_2$ as for elliptic curves makes the smooth points of the curve (i.e., $P \neq (0, 0)$, but including the point at infinity) into an abelian group. Show that the map which takes $P = (x, y)$ to x/y (and takes the point at infinity to zero) gives an isomorphism with the additive group of complex numbers. This is called "additive degeneracy" of an elliptic curve. One way to think of this is to imagine both ω_1 and ω_2 approaching infinity (in different directions). Then g_2 and g_3 both approach zero, so the equation of the corresponding elliptic curve approaches $y^2 = 4x^3$. Meanwhile, the additive group $\mathbb{C}/L$, where $L = \{m\omega_1 + n\omega_2\}$, approaches the additive group $\mathbb{C}$, i.e., the fundamental parallelogram becomes all of $\mathbb{C}$.

11. Let $a \to 0$ in the elliptic curve $y^2 = (x^2 - a)(x + 1)$. Show that for the curve $y^2 = x^2(x + 1)$ the same geometric procedure for finding $P_1 + P_2$ as for elliptic curves makes the smooth points of the curve into an abelian group. Show that the map which takes $P = (x, y)$ to $(y - x)/(y + x)$ (and takes the point at infinity to 1) gives an isomorphism with the multiplicative group $\mathbb{C}^*$ of nonzero complex numbers. This is called "multiplicative degeneracy" of an elliptic curve. Draw the graph of the real points of $y^2 = x^2(x + 1)$, and show where the various sections go under the isomorphism with $\mathbb{C}^*$. One way to think of multiplicative degeneracy is to make the linear change of variables $y \mapsto \frac{1}{2}y$, $x \mapsto -x - \frac{1}{3}$, so that the equation becomes $y^2 = 4x^3 - \frac{4}{3}x - \frac{8}{27}$ (compare with Problem 8 of §I.6). So we are dealing with the limit as t approaches infinity of the group $\mathbb{C}/\{mit + n\pi\}$, i.e., with the vertical strip $\mathbb{C}/\{n\pi\}$ (rather, a cylinder, since opposite sides are glued together), and this is isomorphic to $\mathbb{C}^*$ under the map $z \mapsto e^{2iz}$.

§8. Points of finite order

In any group, there is a basic distinction between elements of finite order and elements of infinite order. In an abelian group, the set of elements of finite order form a subgroup, called the "torsion subgroup". In the case of the group of points in $\mathbb{P}^2_{\mathbb{C}}$ on the elliptic curve $y^2 = f(x)$, we immediately see that a point $P_z = (x, y)$ has finite order if and only if $Nz \in L$ for some N, i.e., if and only if z is a *rational* linear combination of ω_1 and ω_2. In that case, the least such N (which is the least common denominator of the coefficients of ω_1 and ω_2) is the exact order of P_z. Under the isomorphism

from $\mathbb{R}/\mathbb{Z} \times \mathbb{R}/\mathbb{Z}$ to the elliptic curve given by $(a, b) \mapsto P_{a\omega_1 + b\omega_2}$, it is the image of $\mathbb{Q}/\mathbb{Z} \times \mathbb{Q}/\mathbb{Z}$ which is the torsion subgroup of the elliptic curve.

This situation is the two-dimensional analog of the circle group, whose torsion subgroup is precisely the group of all roots of unity, i.e., all $e^{2\pi iz}$ for $z \in \mathbb{Q}/\mathbb{Z}$. Just as the cyclotomic fields—the field extensions of $\mathbb{Q}$ generated by the roots of unity—are central to algebraic number theory, we would expect that the fields obtained by adjoining the coordinates of points $P = (x, y)$ of order N on an elliptic curve should have interesting special properties. We shall soon see that these coordinates are algebraic (if the coefficients of $f(x)$ are). This analogy between cyclotomic fields and fields formed from points of finite order on elliptic curves is actually much deeper than one might have guessed. In fact, a major area of research in algebraic number theory today consists in finding and proving analogs for such fields of the rich results one has for cyclotomic fields.

Let N be a fixed positive integer. Let $f(x) = ax^3 + bx^2 + cx + d = a(x - e_1)(x - e_2)(x - e_3)$ be a cubic polynomial with coefficients in a field K of characteristic $\neq 2$ and with distinct roots (perhaps in some extension of K). We are interested in describing the coordinates of the points of order N (i.e., exact order a divisor of N) on the elliptic curve $y^2 = f(x)$, where these coordinates may lie in an extension of K. If $N = 2$, the points of order N are the point at infinity 0 and $(e_i, 0)$, $i = 1, 2, 3$. Now suppose that $N > 2$. If N is odd, by a "nontrivial" point of order N we mean a point $P \neq 0$ such that $NP = 0$. If N is even, by a "nontrivial" point of order N we mean a point P such that $NP = 0$ but $2P \neq 0$.

Proposition 13. *Let K' be any field extension of K (not necessarily algebraic), and let $\sigma: K' \to \sigma K'$ be any field isomorphism which leaves fixed all elements of K. Let $P \in \mathbb{P}^2_{K'}$ be a point of exact order N on the elliptic curve $y^2 = f(x)$, where $f(x) \in K[x]$. Then σP has exact order N (where for $P = (x, y, z) \in \mathbb{P}^2_{K'}$ we denote $\sigma P \underset{\text{def}}{=} (\sigma x, \sigma y, \sigma z) \in \mathbb{P}^2_{\sigma K'}$).*

Proof. It follows from the addition formulas that $\sigma P_1 + \sigma P_2 = \sigma(P_1 + P_2)$, and hence $N(\sigma P) = \sigma(NP) = \sigma 0 = 0$ (since $\sigma(0, 1, 0) = (0, 1, 0)$). Hence σP has order N. It must have exact order N, since if $N' \sigma P = 0$, we would have $\sigma(N'P) = 0 = (0, 1, 0)$, and hence $N'P = 0$. This proves the proposition. □

Proposition 14. *In the situation of Proposition 13, with K a subfield of $\mathbb{C}$, let $K_N \subset \mathbb{C}$ denote the field obtained by adjoining to K the x- and y-coordinates of all points of order N. Let K_N^+ denote the field obtained by adjoining just their x-coordinates. Then both K_N and K_N^+ are finite galois extensions of K.*

Proof. In each case K_N and K_N^+, we are adjoining a finite set of complex numbers which are permuted by any automorphism of $\mathbb{C}$ which fixes K. This immediately implies the proposition. □

As an example, if $N = 2$, then $K_2 = K_2^+$ is the splitting field of $f(x)$ over K.

Recall that the group of points of order N on an elliptic curve in $\mathbb{P}^2_{\mathbb{C}}$ is isomorphic to $(\mathbb{Z}/N\mathbb{Z}) \times (\mathbb{Z}/N\mathbb{Z})$. Because any $\sigma \in \mathrm{Gal}(K_N/K)$ respects addition of points, i.e., $\sigma(P_1 + P_2) = \sigma P_1 + \sigma P_2$, it follows that each σ gives an invertible linear map of $(\mathbb{Z}/N\mathbb{Z})^2$ to itself.

If R is any commutative ring, we let $GL_n(R)$ denote the group (under matrix multiplication) of all $n \times n$ invertible matrices with entries in R. Here invertibility of a matrix A is equivalent to $\det A \in R^*$, where R^* is the multiplicative group of invertible elements of the ring. For example:

(1) $GL_1(R) = R^*$;
(2) $GL_2(\mathbb{Z}/N\mathbb{Z}) = \{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) | a, b, c, d \in \mathbb{Z}/N\mathbb{Z}, ad - bc \in (\mathbb{Z}/N\mathbb{Z})^*\}$.

It is easy to construct a natural one-to-one correspondence between invertible linear maps $R^n \to R^n$ and elements of $GL_n(R)$. There is no difference with the more familiar case when R is a field.

In our situation of points of order N on an elliptic curve, we have seen that $\mathrm{Gal}(K_N/K)$ is isomorphic to a subgroup of the group of all invertible linear maps $(\mathbb{Z}/N\mathbb{Z})^2 \to (\mathbb{Z}/N\mathbb{Z})^2$. Thus, any $\sigma \in \mathrm{Gal}(K_N/K)$ corresponds to a matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in GL_2(\mathbb{Z}/N\mathbb{Z})$. The matrix entries can be found by writing

$$\sigma P_{\omega_1/N} = P_{a\omega_1/N + c\omega_2/N}, \qquad \sigma P_{\omega_2/N} = P_{b\omega_1/N + d\omega_2/N}.$$

Notice that this is a direct generalization of the situation with the N-th cyclotomic field $\mathbb{Q}_N \underset{\text{def}}{=} \mathbb{Q}(\sqrt[N]{1})$. Recall that $\mathrm{Gal}(\mathbb{Q}_N/\mathbb{Q}) \approx (\mathbb{Z}/N\mathbb{Z})^* = GL_1(\mathbb{Z}/N\mathbb{Z})$, with the element a which corresponds to σ determined by

$$\sigma(e^{2\pi i/N}) = e^{2\pi i a/N}.$$

But one difference in our two-dimensional case of division points on elliptic curves is that, in general, $\mathrm{Gal}(K_N/K) \to GL_2(\mathbb{Z}/N\mathbb{Z})$ is only an injection, not an isomorphism.

In the case $K \subset \mathbb{C}$, say $K = \mathbb{Q}(g_2, g_3)$, where $y^2 = f(x) = 4x^3 - g_2x - g_3$ is in Weierstrass form, we shall now use the $\wp$-function to determine the polynomial whose roots are the x-coordinates of the points of order N. That is, K_N^+ will be the splitting field of such a polynomial.

We first construct an elliptic function $f_N(z)$ whose zeros are precisely the nonzero values of z such that P_z is a point of order N. We follow the prescription in the proof of Proposition 9 of §I.5. If $u \in \mathbb{C}/L$ is a point of order N, then so is the symmetric point $-u$ (which we denoted u^* when we were thinking in terms of points in a fundamental parallelogram). We consider two cases:

(i) N is odd. Then the points u and $-u$ are always distinct modulo L. In other words, u cannot be $\omega_1/2$, $\omega_2/2$ or $(\omega_1 + \omega_2)/2$ if u has odd order N. We define

$$f_N(z) = N \prod (\wp(z) - \wp(u)), \tag{8.1}$$

where the product is taken over nonzero $u \in \mathbb{C}/L$ such that $Nu \in L$, with one u taken from each pair u, $-u$. Then $f_N(z) = F_N(\wp(z))$, where $F_N(x) \in \mathbb{C}[x]$ is a polynomial of degree $(N^2 - 1)/2$. The even elliptic function $f_N(z)$ has $N^2 - 1$ simple zeros and a single pole at 0 of order $N^2 - 1$. Its leading term at $z = 0$ is N/z^{N^2-1}.

(ii) N is even. Now let u range over $u \in \mathbb{C}/L$ such that $Nu \in L$ but u is *not* of order 2, i.e., $u \neq 0, \omega_1/2, \omega_2/2, (\omega_1 + \omega_2)/2$. Define $\tilde{f}_N(z)$ by the product in (8.1). Then $\tilde{f}_N(z) = F_N(\wp(z))$, where $F_N(x) \in \mathbb{C}[x]$ is a polynomial of degree $(N^2 - 4)/2$. The even elliptic function $\tilde{f}_N(z)$ has $N^2 - 4$ simple zeros and a single pole at 0 of order $N^2 - 4$. Its leading term at $z = 0$ is N/z^{N^2-4}.

If N is odd, the function $f_N(z)$ has the property that

$$f_N(z)^2 = N^2 \prod_{0 \neq u \in \mathbb{C}/L, Nu \in L} (\wp(z) - \wp(u)).$$

If N is even, then the function $f_N(z) \underset{\text{def}}{=} \frac{1}{2}\wp'(z)\tilde{f}_N(z)$ has the property that

$$\begin{aligned} f_N(z)^2 &= \tfrac{1}{4}\wp'(z)^2 \tilde{f}_N(z)^2 \\ &= N^2(\wp(z) - e_1)(\wp(z) - e_2)(\wp(z) - e_3) \prod_{u \in \mathbb{C}/L, Nu \in L, 2u \notin L} (\wp(z) - \wp(u)) \\ &= N^2 \prod_{0 \neq u \in \mathbb{C}/L, Nu \in L} (\wp(z) - \wp(u)). \end{aligned}$$

We see that a point $(x, y) = (\wp(z), \wp'(z))$ has odd order N if and only if $F_N(x) = 0$. It has even order N if and only if either $y = 0$ (i.e., it is a point of order 2) or else $F_N(x) = 0$.

Because of Propositions 13 and 14, we know that any automorphism of $\mathbb{C}$ fixing $K = \mathbb{Q}(g_2, g_3)$ permutes the roots of F_N. Hence, the coefficients of F_N are in $K = \mathbb{Q}(g_2, g_3)$.

If we started with an elliptic curve not in Weierstrass form, say $y^2 = f(x) = ax^3 + bx^2 + cx + d$, and if we wanted to avoid using the $\wp$-function, then we could repeatedly apply the addition formulas (7.1)–(7.4) to compute the rational function of x and y which is the x-coordinate of NP, where $P = (x, y)$. We would simplify algebraically as we go, making use of the relation $y^2 = f(x)$, and would end up with an expression in the denominator which vanishes if and only if NP is the point at infinity, i.e., if and only if P has order N (recall: "order N" means "exact order N or a divisor of N").

What type of an expression would we have to get in the denominator of the x-coordinate of NP? Suppose, for example, that N is odd. Then this denominator would be an expression in $K[x, y]$ (with y occurring at most to the first power), where $K = \mathbb{Q}(a, b, c, d)$, which vanishes if and only if x is one of the $(N^2 - 1)/2$ values of x-coordinates of nontrivial points of order N. Thus, the expression must be a polynomial in x alone with $(N^2 - 1)/2$ roots. Similarly, we find that when N is even, this denominator has the form

$y\cdot$(polynomial in x alone), where the polynomial in $K[x]$ has $(N^2-4)/2$ roots.

It is important to note that the algebraic procedure described in the last two paragraphs applies for any elliptic curve $y^2=f(x)$ over any field K of characteristic $\neq 2$, not only over subfields of the complex numbers. Thus, for any K we end up with an expression in the denominator of the x-coordinate of NP that vanishes for at most N^2-1 values of (x, y).

For a general field K, however, we do not necessarily get exactly N^2-1 nontrivial points of order N. Of course, if K is not algebraically closed, the coordinates of points of order N may lie only in some extension of K. Moreover, if K has characteristic p, then there might be fewer points of order N for another reason: the leading coefficient of the expression in the denominator vanishes modulo p, and so the degree of that polynomial drops. We shall soon see examples where there are fewer than N^2 points of order N even if we allow coordinates in K^{algcl}.

This discussion has led to the following proposition.

Proposition 15. *Let $y^2=f(x)$ be an elliptic curve over any field K of characteristic not equal to 2. Then there are at most N^2 points of order N over any extension K' of K.*

Now let us turn our attention briefly to the case of K a finite field, in order to illustrate one application of Proposition 15. We shall later return to elliptic curves over finite fields in more detail.

Since there are only finitely many points in $\mathbb{P}^2_{\mathbb{F}_q}$ (namely, q^2+q+1), there are certainly only finitely many $\mathbb{F}_q$-points on an elliptic curve $y^2=f(x)$, where $f(x)\in\mathbb{F}_q[x]$. So the group of $\mathbb{F}_q$-points is a *finite abelian group*.

By the "type" of a finite abelian group, we mean its expression as a product of cyclic groups of prime power order. We list the orders of all of the cyclic groups that appear in the form: 2^{α_2}, 2^{β_2}, 2^{γ_2}, ..., 3^{α_3}, 3^{β_3}, 3^{γ_3}, ..., 5^{α_5}, 5^{β_5}, ... But Proposition 15 implies that only certain types can occur in the case of the group of $\mathbb{F}_q$-points on $y^2=f(x)$. Namely, for each prime l there are at most two l-th power components l^{α_l}, l^{β_l}, since otherwise we would have more than l^2 points of order l. And of course $l^{\alpha_l+\beta_l}$ must equal the power of l dividing the order of the group.

As an example of how this works, let us consider the elliptic curve $y^2=x^3-n^2x$ over $K=\mathbb{F}_q$ (the finite field of $q=p^f$ elements), where we must assume that p does not divide $2n$. In the case when $q\equiv 3 \pmod 4$, it is particularly easy to count the number of $\mathbb{F}_q$-points.

Proposition 16. *Let $q=p^f$, $p\nmid 2n$. Suppose that $q\equiv 3 \pmod 4$. Then there are $q+1$ $\mathbb{F}_q$-points on the elliptic curve $y^2=x^3-n^2x$.*

PROOF. First, there are four points of order 2: the point at infinity, $(0, 0)$, and $(\pm n, 0)$. We now count all pairs (x, y) where $x\neq 0, n, -n$. We arrange these $q-3$ x's in pairs $\{x, -x\}$. Since $f(x)=x^3-n^2x$ is an odd function, and -1 is not a square in $\mathbb{F}_q$ (here's where we use the assumption that $q\equiv 3 \pmod 4$), it follows that exactly one of the two elements $f(x)$ and $f(-x)=-f(x)$ is a square in $\mathbb{F}_q$. (Recall: In the multiplicative group of a finite field, the squares are a subgroup of index 2, and so the product of two nonsquares is a square, while the product of a square and a nonsquare is a nonsquare.) Whichever of the pair x, $-x$ gives a square, we obtain exactly two points $(x, \pm\sqrt{f(x)})$ or else $(-x, \pm\sqrt{f(-x)})$. Thus, the $(q-3)/2$ pairs give us $q-3$ points. Along with the four points of order two, we have $q+1$ $\mathbb{F}_q$-points in all, as claimed. □

Notice that when $q\equiv 3 \pmod 4$, the number of $\mathbb{F}_q$-points on the elliptic curve $y^2=x^3-n^2x$ does not depend on n. This is not true if $q\equiv 1 \pmod 4$.

As an example, Proposition 16 tells us that for $q=7^3$ there are $344=2^3\cdot 43$ points. Since there are four points of order two, the type of the group of $\mathbb{F}_{343}$-points on $y^2=x^3-n^2x$ must be $(2, 2^2, 43)$.

As a more interesting example, let $q=p=107$. Then there are $108=2^2\cdot 3^3$ points. The group is either of type $(2, 2, 3^3)$ or of type $(2, 2, 3, 3^2)$. To resolve the question, we must determine whether there are 3 or 9 points of order three. (There must be nontrivial points of order 3, since 3 divides the order of the group.) Recall the equation for the x-coordinates of points of order three (see Problem 4 of §7): $-3x^4+6n^2x^2+n^4=0$, i.e., $x=\pm n\sqrt{1\pm 2\sqrt{3}/3}$. Then the corresponding y-coordinates are found by taking $\pm\sqrt{f(x)}$. We want to know how many of these points have both coordinates in $\mathbb{F}_{107}$, rather than an extension of $\mathbb{F}_{107}$. We could compute explicitly, using $\sqrt{3}=\pm 18$ in $\mathbb{F}_{107}$, so that $x=\pm\sqrt{13}, \pm\sqrt{-11}$, etc. But even before doing those computations, we can see that not all 9 points have coordinates in $\mathbb{F}_{107}$. This is because, if (x, y) is in $\mathbb{F}_{107}$, then $(-x, \sqrt{-1}y)$ is another point of order three, and its coordinates are not in $\mathbb{F}_{107}$. Thus, there are only 3 points of order three, and the type of the group is $(2, 2, 3^3)$.

Notice that if K is any field of characteristic 3, then the group of K-points has *no* nontrivial point of order three, because $-3x^4+6n^2x^2+n^4=n^4\neq 0$. This is an example of the "dropping degree" phenomenon mentioned above. It turns out that the same is true for any $p\equiv 3 \pmod 4$, namely, there are no points of order p over a field of characteristic p in that case. This is related to the fact that such p remain prime in the ring of Gaussian integers $\mathbb{Z}[i]$, a ring which is intimately related to our particular elliptic curve (see Problem 13 of §6). But we will not go further into that now.

PROBLEMS

1. For the elliptic curve $y^2=4x^3-g_2x-g_3$, express $\wp(Nz)$ as a rational function of $\wp(z)$ when $N=2$.
2. Let $f_N(z)$ be the elliptic functions defined above. Express $f_3(z)$ as a polynomial in $\wp(z)$.

3. Set $f_1(z)=1$. Prove that for $N=2, 3, 4, \ldots$ we have:
$$\wp(Nz)=\wp(z)-f_{N-1}(z)f_{N+1}(z)/f_N(z)^2.$$
4. In the notation of Proposition 14, suppose that $\sigma\in\text{Gal}(K_N/K)$ fixes all x-coordinates of points of order N. That is, $\sigma|_{K_N^+}=$ identity. Show that the image of σ in $GL_2(\mathbb{Z}/N\mathbb{Z})$ is ± 1. Conclude that $\text{Gal}(K_N/K_N^+)=\{\pm 1\}\cap G$, where G is the image of $\text{Gal}(K_N/K)$ in $GL_2(\mathbb{Z}/N\mathbb{Z})$. What is the analogous situation for cyclotomic fields?
5. Let $L=\{m\omega_1+n\omega_2\}$, and let E be the elliptic curve $y^2=4x^3-g_2(L)x-g_3(L)$. Notice that E does not change if we replace the basis $\{\omega_1, \omega_2\}$ of L by another basis $\{\omega_1', \omega_2'\}$. However, the group isomorphism $\mathbb{C}/L\approx\mathbb{R}/\mathbb{Z}\times\mathbb{R}/\mathbb{Z}$ changes, and so does the isomorphism from the points of order N on E to $\mathbb{Z}/N\mathbb{Z}\times\mathbb{Z}/N\mathbb{Z}$. For example, the point $(\wp(\omega_1'/N), \wp'(\omega_1'/N))$, rather than $(\wp(\omega_1/N), \wp'(\omega_1/N))$, corresponds to $(1, 0)\in\mathbb{Z}/N\mathbb{Z}\times\mathbb{Z}/N\mathbb{Z}$. What effect does the change of basis from ω_i to ω_i' have on the image of $\text{Gal}(K_N/K)$ in $GL_2(\mathbb{Z}/N\mathbb{Z})$?
6. Show that the group $GL_2(\mathbb{Z}/2\mathbb{Z})$ is isomorphic to S_3, the group of permutations of $\{1, 2, 3\}$. For each of the following elliptic curves, describe the image in $GL_2(\mathbb{Z}/2\mathbb{Z})$ of the galois group over $\mathbb{Q}$ of the field generated by the coordinates of the points of order 2.
   (a) $y^2=x^3-nx$ (n not a perfect square)
   (b) $y^2=x^3-n^2x$
   (c) $y^2=x^3-n$ (n not a perfect cube)
   (d) $y^2=x^3-n^3$.
7. (a) How many elements are in $GL_2(\mathbb{Z}/3\mathbb{Z})$?
   (b) Describe the field extension K_3 of $K=\mathbb{Q}$ generated by the coordinates of all points of order 3 on the elliptic curve $y^2=x^3-n^2x$.
   (c) Find $[K_3:\mathbb{Q}]$. What subgroup of $GL_2(\mathbb{Z}/3\mathbb{Z})$ is isomorphic to $\text{Gal}(K_3/\mathbb{Q})$?
   (d) Give a simple example of an element in $GL_2(\mathbb{Z}/3\mathbb{Z})$ that is *not* in the image of $\text{Gal}(K_3/\mathbb{Q})$; in other words, find a pair of elements $z_1=(m_1\omega_1+n_1\omega_2)/3$, $z_2=(m_2\omega_1+n_2\omega_2)/3$ which generate all $(m\omega_1+n\omega_2)/3$ but such that P_{z_1}, P_{z_2} cannot be obtained from $P_{\omega_1/3}$, $P_{\omega_2/3}$ by applying an automorphism to the coordinates of the latter pair of points.
8. In Problem 13 of §I.6, we saw that the lattice corresponding to the curve $y^2=x^3-n^2x$ is the lattice L of Gaussian integers expanded by a factor $\omega_2\in\mathbb{R}$: $L=\{mi\omega_2+n\omega_2\}=\omega_2\mathbb{Z}[i]$.
   (a) Show that the map $z\mapsto iz$ gives an analytic automorphism of the additive group $\mathbb{C}/L$; and, more generally, for any Gaussian integer $a+bi\in\mathbb{Z}[i]$ we have a corresponding analytic endomorphism of $\mathbb{C}/L$ induced by $z\mapsto(a+bi)z$.
   (b) Notice that if $b=0$, this is the map $z\mapsto z+z+\cdots+z$ (a times) which gives ϕ_a: $P\mapsto aP$ on the elliptic curve. By looking at the definition of $\wp(z)$, $\wp'(z)$, show that the map $z\mapsto iz$ gives the automorphism ϕ_i: $(x, y)\mapsto(-x, iy)$ on the elliptic curve. This is an example of what's called "complex multiplication". Show that $\phi_i\circ\phi_i=\phi_{-1}$, and in fact the map $a+bi\mapsto\phi_{a+bi}$ is an injection of the ring $\mathbb{Z}[i]$ into the ring of endomorphisms of the group of points on the elliptic curve.
   (c) If L is a lattice in $\mathbb{C}$ and if there exists a complex number $\alpha=a+bi$, $b\neq 0$, such that $\alpha L\subset L$, show that α is a quadratic imaginary algebraic integer, and that L contains a sublattice of finite index of the form $\omega_2\mathbb{Z}[\alpha]$.

9. Each of the following points has finite order N on the given elliptic curve. In each case, find its order.
   (a) $P=(0, 4)$ on $y^2=4x^3+16$
   (b) $P=(2, 8)$ on $y^2=4x^3+16x$
   (c) $P=(2, 3)$ on $y^2=x^3+1$
   (d) $P=(3, 8)$ on $y^2=x^3-43x+166$
   (e) $P=(3, 12)$ on $y^2=x^3-14x^2+81x$
   (f) $P=(0, 0)$ on $y^2+y=x^3-x^2$
   (g) $P=(1, 0)$ on $y^2+xy+y=x^3-x^2-3x+3$.

§9. Points over finite fields, and the congruent number problem

We have mainly been interested in elliptic curves E over $\mathbb{Q}$, particularly the elliptic curve $y^2=x^3-n^2x$, which we shall denote E_n. But if K is any field whose characteristic p does not divide $2n$, the same equation (where we consider n modulo p) is an elliptic curve over K. We shall let $E_n(K)$ denote the set of points on the curve with coordinates in K. Thus, Proposition 16 in the last section can be stated: If $q\equiv 3 \pmod 4$, then $\#E_n(\mathbb{F}_q)=q+1$.

The elliptic curve E_n considered as being defined over $\mathbb{F}_p$, is called the "reduction" modulo p, and we say that E_n has "good reduction" if p does not divide $2n$, i.e., if $y^2=x^3-n^2x$ gives an elliptic curve over $\mathbb{F}_p$. More generally, if $y^2=f(x)$ is an elliptic curve E defined over an algebraic number field, and if $\mathfrak{p}$ is a prime ideal of the number field which does not divide the denominators of the coefficients of $f(x)$ or the discriminant of $f(x)$, then by reduction modulo $\mathfrak{p}$ we obtain an elliptic curve defined over the (finite) residue field of $\mathfrak{p}$.

At first glance, it may seem that the elliptic curves over finite fields—which lead only to finite abelian groups—are not a serious business, and that reduction modulo p is a frivolous game that will not help us in our original objective of studying $\mathbb{Q}$-points on $y^2=x^3-n^2x$. However, this is far from the case. Often information from the various reductions modulo p can be pieced together to yield information about the $\mathbb{Q}$-points. This is usually a subtle and difficult procedure, replete with conjectures and unsolved problems. However, there is one result of this type which is simple enough to give right now. Namely, we shall use reduction modulo p for various primes p to determine the torsion subgroup of $E_n(\mathbb{Q})$, the group of $\mathbb{Q}$-points on $y^2=x^3-n^2x$.

In any abelian group, the elements of finite order form a subgroup, called the "torsion subgroup". For example, the group $E(\mathbb{C})$ of complex points on an elliptic curve is isomorphic to $\mathbb{C}/L$, which for any lattice L is isomorphic to $\mathbb{R}/\mathbb{Z}\times\mathbb{R}/\mathbb{Z}$ (see Problem 2 of §I.5). Its torsion subgroup corresponds to the subgroup $\mathbb{Q}/\mathbb{Z}\times\mathbb{Q}/\mathbb{Z}\subset\mathbb{R}/\mathbb{Z}\times\mathbb{R}/\mathbb{Z}$, i.e., in $\mathbb{C}/L$ it consists of all rational linear combinations of ω_1 and ω_2.

A basic theorem of Mordell states that the group $E(\mathbb{Q})$ of $\mathbb{Q}$-points on an

elliptic curve E defined over $\mathbb{Q}$ is a finitely generated abelian group. This means that (1) the torsion subgroup $E(\mathbb{Q})_{\text{tors}}$ is finite, and (2) $E(\mathbb{Q})$ is isomorphic to the direct sum of $E(\mathbb{Q})_{\text{tors}}$ and a finite number of copies of $\mathbb{Z}$: $E(\mathbb{Q}) \approx E(\mathbb{Q})_{\text{tors}} \oplus \mathbb{Z}^r$. The nonnegative integer r is called the "rank" of $E(\mathbb{Q})$. It is greater than zero if and only if E has infinitely many $\mathbb{Q}$-points. Mordell's theorem is also true, by the way, if $\mathbb{Q}$ is replaced by any algebraic number field. This generalization, proved by Andre Weil, is known as the Mordell–Weil theorem. We shall not need this theorem for our purposes, even in the form proved by Mordell. For a proof, the reader is referred to Husemuller's forthcoming book on elliptic curves or else to [Lang 1978b].

We shall now prove that the only rational points of finite order on E_n are the four points of order 2: 0 (the point at infinity), $(0, 0)$, $(\pm n, 0)$.

Proposition 17. $\#E_n(\mathbb{Q})_{\text{tors}} = 4$.

PROOF. The idea of the proof is to construct a group homomorphism from $E_n(\mathbb{Q})_{\text{tors}}$ to $E_n(\mathbb{F}_p)$ which is injective for most p. That will imply that the order of $E_n(\mathbb{Q})_{\text{tors}}$ divides the order of $E_n(\mathbb{F}_p)$ for such p. But no number greater than 4 could divide all such numbers $\#E_n(\mathbb{F}_p)$, because we at least know that $\#E_n(\mathbb{F}_p)$ runs through all integers of the form $p+1$ for p a prime congruent to 3 modulo 4 (see Proposition 16).

We begin the proof of Proposition 17 by constructing the homomorphism from the group of $\mathbb{Q}$-points on E_n to the group of $\mathbb{F}_p$-points. More generally, we simply construct a map from $\mathbb{P}^2_{\mathbb{Q}}$ to $\mathbb{P}^2_{\mathbb{F}_p}$. In what follows, we shall always choose a triple (x, y, z) for a point in $\mathbb{P}^2_{\mathbb{Q}}$ in such a way that x, y, and z are integers with no common factor. Up to multiplication by ± 1, there is a unique such triple in the equivalence class. For any fixed prime p, we define the image $\bar{P}$ of $P = (x, y, z) \in \mathbb{P}^2_{\mathbb{Q}}$ to be the point $\bar{P} = (\bar{x}, \bar{y}, \bar{z}) \in \mathbb{P}^2_{\mathbb{F}_p}$, where the bar denotes reduction of an integer modulo p. Note that $\bar{P}$ is not the identically zero triple, because p does not divide all three integers x, y, z. Also note that we could have replaced the triple (x, y, z) by any multiple by an integer prime to p without affecting $\bar{P}$.

It is easy to see that if $P = (x, y, z)$ happens to be in $E_n(\mathbb{Q})$, i.e., if $y^2z = x^3 - n^2xz^2$, then $\bar{P}$ is in $E_n(\mathbb{F}_p)$. Moreover, the image of $P_1 + P_2$ under this map is $\bar{P}_1 + \bar{P}_2$, because it makes no difference whether we use the addition formulas (7.1)–(7.4) to find the sum and then reduce mod p, or whether we first reduce mod p and then use the addition formulas. In other words, our map is a homomorphism from $E_n(\mathbb{Q})$ to $E_n(\mathbb{F}_p)$, for any prime p not dividing $2n$.

We now determine when this map is not injective, i.e., when two points $P_1 = (x_1, y_1, z_1)$ and $P_2 = (x_2, y_2, z_2)$ in $\mathbb{P}^2_{\mathbb{Q}}$ have the same image $\bar{P}_1 = \bar{P}_2$ in $\mathbb{P}^2_{\mathbb{F}_p}$.

Lemma. *$\bar{P}_1 = \bar{P}_2$ if and only if the cross-product of P_1 and P_2 (considered as vectors in $\mathbb{R}^3$) is divisible by p, i.e., if and only if p divides $y_1z_2 - y_2z_1$, $x_2z_1 - x_1z_2$, and $x_1y_2 - x_2y_1$.*

PROOF OF LEMMA. First suppose that p divides the cross-product. We consider two cases:

(i) p divides x_1. Then p divides x_2z_1 and x_2y_1, and therefore divides x_2, because it cannot divide x_1, y_1 and z_1. Suppose, for example, that $p \nmid y_1$ (an analogous argument will apply if $p \nmid z_1$). Then $\bar{P}_2 = (0, \bar{y}_1\bar{y}_2, \bar{y}_1\bar{z}_2) = (0, \bar{y}_1\bar{y}_2, \bar{y}_2\bar{z}_1) = (0, \bar{y}_1, \bar{z}_1) = \bar{P}_1$ (where we have used the fact that p divides $y_1z_2 - y_2z_1$).

(ii) p does not divide x_1. Then $\bar{P}_2 = (\bar{x}_1\bar{x}_2, \bar{x}_1\bar{y}_2, \bar{x}_1\bar{z}_2) = (\bar{x}_1\bar{x}_2, \bar{x}_2\bar{y}_1, \bar{x}_2\bar{z}_1) = (\bar{x}_1, \bar{y}_1, \bar{z}_1) = \bar{P}_1$.

Conversely, suppose that $\bar{P}_1 = \bar{P}_2$. Without loss of generality, suppose that $p \nmid x_1$ (an analogous argument will apply if $p \nmid y_1$ or $p \nmid z_1$). Then, since $\bar{P}_1 = \bar{P}_2 = (\bar{x}_2, \bar{y}_2, \bar{z}_2)$, we also have $p \nmid x_2$. Hence, $(\bar{x}_1\bar{x}_2, \bar{x}_1\bar{y}_2, \bar{x}_1\bar{z}_2) = \bar{P}_2 = \bar{P}_1 = (\bar{x}_2\bar{x}_1, \bar{x}_2\bar{y}_1, \bar{x}_2\bar{z}_1)$. Since the first coordinates are the same, these two points can be equal only if the second and third coordinates are equal, i.e., if p divides $x_1y_2 - x_2y_1$ and $x_1z_2 - x_2z_1$. Finally, we must show that p divides $y_1z_2 - y_2z_1$. If both y_1 and z_1 are divisible by p, then this is trivial. Otherwise, the conclusion will follow by repeating the above argument with x_1, x_2 replaced by y_1, y_2 or by z_1, z_2. This concludes the proof of the lemma.

We are now ready to prove Proposition 17. Suppose that the proposition is false, i.e., that $E_n(\mathbb{Q})$ contains a point of finite order greater than 2. Then either it contains an element of odd order, or else the group of points of order 4 (or a divisor of 4) contains either 8 or 16 elements. In either case we have a subgroup $S = \{P_1, P_2, \ldots, P_m\} \subset E_n(\mathbb{Q})_{\text{tors}}$, where $m = \#S$ is either 8 or else an odd number.

Let us write all of the points P_i, $i = 1, \ldots, m$, in the form in the lemma: $P_i = (x_i, y_i, z_i)$. For each pair of points P_i, P_j, consider the cross-product vector $(y_iz_j - y_jz_i, x_jz_i - x_iz_j, x_iy_j - x_jy_i) \in \mathbb{R}^3$. Since P_i and P_j are distinct points, as vectors in $\mathbb{R}^3$ they are not proportional, and so their cross-product is not the zero vector. Let n_{ij} be the greatest common divisor of the coordinates of this cross-product. According to the lemma, the points P_i and P_j have the same image $\bar{P}_i = \bar{P}_j$ in $E_n(\mathbb{F}_p)$ if and only if p divides n_{ij}. Thus, if p is a prime of good reduction which is greater than all of the n_{ij}, it follows that all images are distinct, i.e., the map reduction modulo p gives an *injection* of S in $E_n(\mathbb{F}_p)$.

But this means that for all but finitely many p the number m must divide $\#E_n(\mathbb{F}_p)$, because the image of S is a subgroup of order m. Then for all but finitely many primes congruent to 3 modulo 4, by Proposition 16 we must have $p \equiv -1 \pmod{m}$. But this contradicts Dirichlet's theorem on primes in an arithmetic progression. Namely, if $m = 8$ this would mean that there are only finitely many primes of the form $8k + 3$. If m is odd, it would mean that there are only finitely many primes of the form $4mk + 3$ (if $3 \nmid m$), and that there are only finitely many primes of the form $12k + 7$ if $3 | m$. In all cases, Dirichlet's theorem tells us that there are infinitely many primes of the given type. This concludes the proof of Proposition 17. □

Notice how the technique of reduction modulo p (more precisely, the use of Proposition 16 for infinitely many primes p) led to a rather painless proof of a strong fact: There are no "non-obvious" rational points of finite order on E_n. As we shall soon see, this fact is useful for the congruent number problem. But a far more interesting and difficult question is the existence of points of infinite order, i.e., whether the rank r of $E_n(\mathbb{Q})$ is nonzero. As we shall see in a moment, that question is actually *equivalent* to the question of whether or not n is a congruent number.

So it is natural to ask whether mod p information can somehow be put together to yield information about the rank of an elliptic curve. This subtle question will lead us in later chapters to consideration of the Birch–Swinnerton–Dyer conjecture for elliptic curves.

For further general motivational discussion of elliptic curves over finite fields, see [Koblitz 1982].

We now prove the promised corollary of Proposition 17.

Proposition 18. *n is a congruent number if and only if $E_n(\mathbb{Q})$ has nonzero rank r.*

PROOF. First suppose that n is a congruent number. At the beginning of §2, we saw that the existence of a right triangle with rational sides and area n leads to a rational point on E_n whose x-coordinate lies in $(\mathbb{Q}^+)^2$. Since the x-coordinates of the three nontrivial points of order 2 are $0, \pm n$, this means that there must be a rational point not of order 2. By Proposition 17, such a point has infinite order, i.e., $r \geq 1$.

Conversely, suppose that P is a point of infinite order. By Problem 2(c) of §I.7, the x-coordinate of the point $2P$ is the square of a rational number having even denominator. Now by Proposition 2 in §I.2, the point $2P$ corresponds to a right triangle with rational sides and area n (under the correspondence in Proposition 1). This proves Proposition 18. □

Notice the role of Proposition 17 in the proof of Proposition 18. It tells us that the only way to get nontrivial rational points of the form $2P$ is from points of infinite order. Let $2E_n(\mathbb{Q})$ denote the subgroup of $E_n(\mathbb{Q})$ consisting of the doubles of rational points. Then Proposition 17 is equivalent to the assertion that $2E_n(\mathbb{Q})$ is a torsion-free abelian group, i.e., it is isomorphic to a certain number of copies (namely, r) of $\mathbb{Z}$. The set $2E_n(\mathbb{Q}) - 0$ (0 denotes the point at infinity) is empty if and only if $r = 0$.

We saw that points in the set $2E_n(\mathbb{Q}) - 0$ lead to right triangles with rational sides and area n under the correspondence in Proposition 1. It is natural to ask whether all points meeting the conditions in Proposition 2, i.e., corresponding to triangles, are doubles of points. We now prove that the answer is yes. At the same time, we give another verification of Proposition 18 (not relying on the homework problem 2(c) of §I.7).

Proposition 19. *There is a one-to-one correspondence between right triangles with rational sides $X < Y < Z$ and area n, and pairs of points $(x, \pm y) \in 2E_n(\mathbb{Q}) - 0$. The correspondence is:*

$$(x, \pm y) \mapsto \sqrt{x+n} - \sqrt{x-n},\ \sqrt{x+n} + \sqrt{x-n},\ 2\sqrt{x};$$
$$X, Y, Z \mapsto (Z^2/4,\ \pm(Y^2 - X^2)Z/8).$$

In light of Proposition 1 of §I.1, Proposition 19 is an immediate consequence of the following general characterization of the doubles of points on elliptic curves.

Proposition 20. *Let E be the elliptic curve $y^2 = (x - e_1)(x - e_2)(x - e_3)$ with $e_1, e_2, e_3 \in \mathbb{Q}$. Let $P = (x_0, y_0) \in E(\mathbb{Q}) - 0$. Then $P \in 2E(\mathbb{Q}) - 0$ if and only if $x_0 - e_1$, $x_0 - e_2$, $x_0 - e_3$ are all squares of rational numbers.*

PROOF. We first note that, without loss of generality, we may assume that $x_0 = 0$. To see this, make the change of variables $x' = x - x_0$. By simply translating the geometrical picture for adding points, we see that the point $P' = (0, y_0)$ on the curve E' with equation $y^2 = (x - e_1')(x - e_2')(x - e_3')$, where $e_i' = e_i - x_0$, is in $2E'(\mathbb{Q}) - 0$ if and only if our original P were in $2E(\mathbb{Q}) - 0$. And trivially, the $x_0 - e_i$ are all squares if and only if the $(0 - e_i')$ are. So it suffices to prove the proposition with $x_0 = 0$.

Next, note that if there exists $Q \in E(\mathbb{Q})$ such that $2Q = P$, then there are exactly four such points Q, Q_1, Q_2, $Q_3 \in E(\mathbb{Q})$ with $2Q_i = P$. To obtain Q_i, simply add to Q the point of order two $(e_i, 0) \in E(\mathbb{Q})$ (see Problem 5 in §I.7).

Choose a point $Q = (x, y)$ such that $2Q = P = (0, y_0)$. We want to find conditions for the coordinates of one such Q (and hence all four) to be rational. Now a point Q on the elliptic curve satisfies $2Q = P$ if and only if the tangent line to the curve at Q passes through $-P = (0, -y_0)$. That is, the four possible points Q are obtained geometrically by drawing the four distinct lines emanating from $-P$ which are tangent to the curve.

We readily verify that the coordinates (x, y) are rational if and only if the slope of the line from $-P$ to Q is rational. The "only if" is immediate. Conversely, if this slope m is rational, then the x-coordinate of Q, which is the double root of the cubic $(mx - y_0)^2 = (x - e_1)(x - e_2)(x - e_3)$, must also be rational. (Explicitly, $x = (e_1 + e_2 + e_3 + m^2)/2$.) In this case the y-coordinate of Q is also rational: $y = mx - y_0$. Thus, we want to know when one (and hence all four) slopes of lines from $-P$ which are tangent to E are rational.

A number $m \in \mathbb{C}$ is the slope of a line from $-P$ which is tangent to E if and only if the following equation has a double root:

$$(mx - y_0)^2 = (x - e_1)(x - e_2)(x - e_3) = x^3 + ax^2 + bx + c, \quad (9.1)$$

with

$$a = -e_1 - e_2 - e_3, \qquad b = e_1e_2 + e_1e_3 + e_2e_3, \qquad c = -e_1e_2e_3 = y_0^2, \quad (9.2)$$

where the last equality $c = y_0^2$ comes from the fact that $(0, y_0)$ is on the curve

$y^2 = x^3 + ax^2 + bx + c$. Now if we simplify (9.1) and factor out x, our condition becomes: the following quadratic equation has a double root:

$$x^2 + (a - m^2)x + (b + 2my_0) = 0.$$

This is equivalent to saying that its discriminant must vanish, i.e.,

$$(a - m^2)^2 - 4(b + 2my_0) = 0. \tag{9.3}$$

Thus, our task is to determine when one (and hence all four) roots of this quartic polynomial in m are rational.

We want to find a condition in terms of the e_i's (namely, our claim is that an equivalent condition is: $-e_i \in \mathbb{Q}^2$). In (9.3), the a and b are symmetric polynomials in the e_i, but the y_0 is not. However, y_0 is a symmetric polynomial in the $\sqrt{e_i}$. That is, we introduce f_i satisfying $f_i^2 = -e_i$. There are two possible choices for f_i, unless $e_i = 0$. Choose the f_i in any of the possible ways, subject to the condition that $y_0 = f_1 f_2 f_3$. If all of the e_i are nonzero, this means that the sign of f_1 and f_2 are arbitrary, and then the sign of f_3 is chosen so that y_0 and $f_1 f_2 f_3$ are the same square root of $-e_1 e_2 e_3$. If, say, $e_3 = 0$, then either choice can be made for the sign of f_1, f_2, and of course $f_3 = 0$. In all cases there are four possible choices of the f_i's consistent with the requirement that $y_0 = f_1 f_2 f_3$. Once we fix one such choice f_1, f_2, f_3, we can list the four choices as follows (here we're supposing that e_1 and e_2 are nonzero):

$$f_1, f_2, f_3; \qquad f_1, -f_2, -f_3; \qquad -f_1, f_2, -f_3; \qquad -f_1, -f_2, f_3. \tag{9.4}$$

The advantage of going from the e_i's to the f_i's is that now the coefficients of our equation (9.3) are symmetric functions of f_1, f_2, f_3. More precisely, if we set $s_1 = f_1 + f_2 + f_3$, $s_2 = f_1 f_2 + f_1 f_3 + f_2 f_3$, $s_3 = f_1 f_2 f_3$, the elementary symmetric functions, then

$$\begin{aligned} a &= f_1^2 + f_2^2 + f_3^2 = s_1^2 - 2s_2; \\ b &= f_1^2 f_2^2 + f_1^2 f_3^2 + f_2^2 f_3^2 = s_2^2 - 2s_1 s_3; \\ y_0 &= s_3. \end{aligned}$$

Thus, equation (9.3) becomes

$$\begin{aligned} 0 &= (m^2 - s_1^2 + 2s_2)^2 - 4(s_2^2 - 2s_1 s_3 + 2ms_3) \\ &= (m^2 - s_1^2)^2 + 4s_2(m^2 - s_1^2) - 8s_3(m - s_1). \end{aligned} \tag{9.5}$$

We see at a glance that the polynomial in (9.5) is divisible by $m - s_1$, i.e., $m = s_1 = f_1 + f_2 + f_3$ is a root. Since we could have made three other choices for the signs of the f_i, the other roots must correspond to these choices, i.e., the four solutions of equation (9.3) are:

$$\begin{aligned} m_1 &= f_1 + f_2 + f_3, & m_2 &= f_1 - f_2 - f_3, \\ m_3 &= -f_1 + f_2 - f_3, & m_4 &= -f_1 - f_2 + f_3. \end{aligned} \tag{9.6}$$

We want to know whether the four values in (9.6) are rational. Clearly, if all of the f_i are rational, then so are the m_i. Conversely, suppose the m_i are rational. Then $f_1 = (m_1 + m_2)/2$, $f_2 = (m_1 + m_3)/2$, and $f_3 = (m_1 + m_4)/2$ are rational. The conclusion of this string of equivalent conditions is: the coordinates (x, y) of a point Q for which $2Q = P$ are rational if and only if the $f_i = \sqrt{-e_i}$ are rational. This proves Proposition 20. □

Finally, we note that Proposition 20 holds with $\mathbb{Q}$ replaced by any field K not of characteristic 2. Essentially the same proof applies. (We need only take care to use algebraic rather than geometric arguments, for example, when reducing to the case $P = (0, y_0)$.)

Problems

1. Prove that for f odd, any $\mathbb{F}_{p^f}$-point of order 3 on the elliptic curve E_n: $y^2 = x^3 - n^2 x$ is actually an $\mathbb{F}_p$-point; prove that there are at most three such points if $p \equiv 3 \pmod 4$; and find a fairly good sufficient condition on p and f which ensures nine $\mathbb{F}_{p^f}$-points of order 3.

2. For each of the following values of q, find the order and type of the group of $\mathbb{F}_q$-points on the elliptic curve E_1: $y^2 = x^3 - x$. In all cases, find the type directly, if necessary checking how many points have order 3 or 4. Don't "peek" at the later problems.
   (a) All odd primes from 3 to 23.
   (b) 9
   (c) 27
   (d) 71
   (e) 11^3.

3. Find the type of the group of $\mathbb{F}_p$-points on the elliptic curve E_5: $y^2 = x^3 - 25x$ for all odd primes p of good reduction up to 23.

4. Prove that for $a \in \mathbb{Q}$ the equation $y^2 = x^3 - a$ determines an elliptic curve over any field K whose characteristic p does not divide 6 or the numerator or denominator of a; and that it has $q + 1$ $\mathbb{F}_q$-points if $q \equiv 2 \pmod 3$.

5. Prove that there are exactly 3 $\mathbb{F}_q$-points of order 3 on the elliptic curve in Problem 4 if $q \equiv 2 \pmod 3$.

6. For all odd primes p from 5 to 23, find the order and type of the group of $\mathbb{F}_p$-points on the elliptic curve $y^2 = x^3 - 1$.

7. Prove that the torsion subgroup of the group of $\mathbb{Q}$-points on the elliptic curve $y^2 = x^3 - a$ has order at most 6, and that its order is equal to:
   (a) 6 if $a = -b^6$ for some $b \in \mathbb{Q}$;
   (b) 2 if $a = c^3$ for some $c \in \mathbb{Q}$ with c not of the form $-b^2$;
   (c) 3 if either $a = -d^2$ for some $d \in \mathbb{Q}$ with d not of the form b^3, or if $a = 432b^6$ for some $b \in \mathbb{Q}$;
   (d) 1 otherwise.

8. Show that the correspondence constructed in Problem 2 of §I.2 gives a one-to-one correspondence between right triangles as in Proposition 19 and pairs $\pm P$ of

non-identity elements of the quotient group $E_n(\mathbb{Q})/E_n(\mathbb{Q})_{\text{torsion}}$, which is isomorphic to $2E_n(\mathbb{Q})$ under the map $P \mapsto 2P$. See Problem 2(b) of §I.7.

In the problems below, we illustrate how more information can be obtained using two additional tools: (1) the complex multiplication automorphism $(x, y) \mapsto (-x, \sqrt{-1}y)$ of the group of K-points of the elliptic curve $y^2 = x^3 - n^2 x$ if K contains a square root of -1; (2) the action of $\mathrm{Gal}(K^{\mathrm{alg\,cl}}/K)$ on the coordinates of the $K^{\mathrm{alg\,cl}}$-points.

9. Suppose that $q \equiv 3 \pmod 4$, and l is an odd prime. Prove that:
   (a) there are at most l $\mathbb{F}_q$-points of order l on the elliptic curve $y^2 = x^3 - n^2 x$, and there are at most eight $\mathbb{F}_q$-points of order 4;
   (b) the group of $\mathbb{F}_q$-points is the product of a group of order 2 and a cyclic group of order $(q + 1)/2$.

10. Suppose that $q \equiv 2 \pmod 3$, $2 \nmid N$, $3 \nmid N$. Prove that there are at most N $\mathbb{F}_q$-points of order N on the elliptic curve $y^2 = x^3 - a$.

11. Suppose that $q \equiv 1 \pmod 4$, and $l \equiv 3 \pmod 4$ is a prime not equal to p. Let (l^α, l^β) be the l-part of the type of the group of $\mathbb{F}_q$-points on the elliptic curve $y^2 = x^3 - n^2 x$. Prove that $\alpha = \beta$. If $l = 2$, prove that $\alpha = \beta$ or $\alpha = \beta \pm 1$.

12. The group of K-points on an elliptic curve is analogous to the multiplicative group K^*. In Problem 11 of §I.7, we saw that for $K = \mathbb{C}$, as $a \to 0$ the elliptic curve $y^2 = (x^2 - a)(x + 1)$ "becomes" the multiplicative group $\mathbb{C}^*$. Now let K be the finite field $\mathbb{F}_q$. In this problem we work with K^*, and in the next problem we work with the group of K-points on an elliptic curve. Let l be a prime not equal to p, and suppose that $\mathbb{F}_q$ contains all l-th roots of 1, i.e., $q = p^f \equiv 1 \pmod l$.
   (a) Show that the splitting field of $x^l - a$, where $a \in \mathbb{F}_q$, has degree either 1 or l over $\mathbb{F}_q$.
   (b) Show that the subfield of $\mathbb{F}_q^{\mathrm{alg\,cl}}$ generated by all l^{M+1}-th roots of 1 is $\mathbb{F}_{q^{l^{M'}}}$, where $M' \le M$.
   (c) (For readers who know about l-adic numbers.) Construct an isomorphism between the additive group $\mathbb{Z}_l$ of l-adic integers and the galois group over $\mathbb{F}_q$ of the field extension generated by all l-th power division points (i.e., l-th power roots of unity).

13. Now let E be an elliptic curve defined over $\mathbb{F}_q$. Suppose that there are l^2 $\mathbb{F}_q$-points of order l.
   (a) Let A be an $\mathbb{F}_q$-point, and let $\mathbb{F}_{q^r}$ be the extension of $\mathbb{F}_q$ generated by the coordinates of a solution α to the equation $l\alpha = A$ (i.e., $\mathbb{F}_{q^r}$ is the smallest extension of $\mathbb{F}_q$ containing such an α). Show that there are l^2 $\mathbb{F}_{q^r}$-points α_i such that $l\alpha_i = A$.
   (b) Fix an $\mathbb{F}_{q^r}$-point α such that $l\alpha = A$. Prove that the map $\sigma \mapsto \sigma(\alpha) - \alpha$ gives an imbedding of $\mathrm{Gal}(\mathbb{F}_{q^r}/\mathbb{F}_q)$ into the group of points of order l on E.
   (c) Show that $r = 1$ or l.
   (d) What is the field extension of $\mathbb{F}_q$ generated by all points of order l^M, $M = 1, 2, \ldots$? What is its galois group?

CHAPTER II

The Hasse–Weil L-Function of an Elliptic Curve

At the end of the last chapter, we used reduction modulo p to find some useful information about the elliptic curves E_n: $y^2 = x^3 - n^2 x$ and the congruent number problem. We considered E_n as a curve over the prime field $\mathbb{F}_p$ where $p \nmid 2n$; used the easily proved equality $\#E_n(\mathbb{F}_p) = p + 1$ when $p \equiv 3 \pmod 4$; and, by making use of infinitely many such p, were able to conclude that the only rational points of finite order on E_n are the four obvious points of order two. This then reduced the congruent number problem to the determination of whether r, the rank of $E_n(\mathbb{Q})$, is zero or greater than zero.

Determining r is much more difficult than finding the torsion group. Some progress can be made using the number of $\mathbb{F}_q$-points. But the progress does not come cheaply. First of all, we will derive a formula for $\#E_n(\mathbb{F}_q)$ for any prime power $q = p^r$. Next, we will combine these numbers $N_r = N_{r,p} = \#E_n(\mathbb{F}_{p^r})$ into a function which is analogous to the Riemann zeta-function (but more complicated). The behavior of this complex-analytic function near the point 1 is intimately related to the group of rational points.

Before introducing this complex-analytic function, which is defined using all of the $N_{r,p}$, we introduce a much simpler function, called the "congruence zeta-function", which is built up from the $N_r = N_{r,p}$ for a fixed prime p.

§1. The congruence zeta-function

Given any sequence N_r, $r = 1, 2, 3, \ldots$, we define the corresponding "zeta-function" by the formal power series

$$Z(T) = \exp\left(\sum_{r=1}^{\infty} N_r \frac{T^r}{r}\right) \qquad \text{where} \quad \exp(u) \overset{\text{def}}{=} \sum_{k=0}^{\infty} \frac{u^k}{k!}. \tag{1.1}$$