CRD Expository Report 31
CCR Expository 34
An Elementary Introduction to Elliptic Curves I and II

Leonard S. Charlap
David P. Robbins
Raymond Coley

minor additions and corrections S. Kunz-Jacques

December 1988 - July 1990

Abstract

In his paper on elliptic curves over finite fields, R. Schoof assumes certain basic material concerning elliptic curves. This material mainly concerns the division polynomials and the "Weil Conjectures for Elliptic Curves". The two first chapters of the present report provide elementary self-contained proofs of these results. Chapter 3 is concerned with rational maps between elliptic curves and Weil reciprocity. We prove that all isogenies are homomorphisms and the "Lower Star Theorem", as well as generalized Weil reciprocity.

Keywords: conjecture, curve, division, elliptic, Frobenius, function, isogeny, map, pair, polynomial, ramification, rational, reciprocity, Reimann, Roch, uniformizing, Weil

Chapter 1

Elliptic Curves over Algebraically Closed Fields

1.1 Introduction

The goal of these notes is to prove the results used by Schoof in his paper on elliptic curves over finite fields. On the way, we expose most of the basic notions of elliptic curve theory required for further study. It appears to be impossible to find an elementary presentation of this material in the literature. By "elementary", we mean that the exposition requires little beyond undergraduate mathematics. It is still true, however, that our "elementary" proofs may require some mathematical sophistication. Thus you would not have to consult a lot of other books (as in [6]), but you still may have to expend some thought.

We would like to thank our colleagues at IDA for their unrelenting criticism and good advice. In addition, we would like to thank

Ann Stehney for her editorial and mathematical suggestions.

Recall that the *characteristic* of a field K is the smallest positive integer p such that $p \cdot 1 = 0$ where by $p \cdot 1$ we mean $1 + 1 + \cdots + 1$, p times. It can be easily seen that if there is such a p, it is always a prime integer (see [4], page 241). If there is no such positive integer p, we say the characteristic is zero.

Recall that a field K is said to be *algebraically closed* if every polynomial with coefficients in K splits completely into linear factors. Another way of saying this is that every polynomial of degree n (with coefficients in K) has n roots in K, counting multiplicities. The standard example of an algebraically closed field is $\mathbb{C}$, the complex numbers. It is a standard fact (see [2], page 107, for example) that every field has an algebraic closure, *i.e.*, an algebraically closed superfield.

We assume that the characteristic of our field K is not 2 and that K is algebraically closed.

We will use some standard notation that we record here. $K[X]$ will denote the ring of polynomials in the indeterminate X with coefficients in K while $K(X)$ will denote its field of quotients, namely the field of rational functions in X with coefficients in K. $K[X, Y]$ and $K(X, Y)$ are defined similarly.

1.2 Elliptic Curves

We give the basic definition.

Definition 1.2.1. For any $A, B \in K$, we can define an *elliptic curve E*. E is the set of all points $(h, k) \in K \times K$ that satisfy the equation

$$k^2 = h^3 + Ah + B \tag{1.1}$$

together with an "idealized point" $\mathcal{O}$. For reasons that will become apparent later, $\mathcal{O}$ is called the *identity*. The points of the curve other than the identity are said to be *finite*.

Definition 1.2.2. If k is a subfield of K, and A and B are in k, we define the *k-rational* points of E to be the points whose coordinates lie in k. We denote the set of these points by $E(k)$.

Definition 1.2.3. An elliptic curve defined by Equation (1.1) is called *nonsingular* if the polynomial $X^3 + AX + B$ has (three) distinct roots; otherwise we say it is *singular*.

Exercise 1.2.4. Define $\Delta(E) = 4A^3 + 27B^2 . \Delta(E)$ is called the *discriminant* of the equation of the curve. Show that E is nonsingular if and only if $\Delta(E) \neq 0$.

Remarks.

(i) The symbols x and y will be reserved for the coordinate functions on E defined by $x(a, b) = a$ and $y(a, b) = b$.

(ii) It is sometimes fruitful to think of the identity as being "at infinity." If we use projective coordinates, we can actually make sense of this notion (see [6]). We will abuse the terminology and use the symbol ∞ for both the x and the y coordinates of $\mathcal{O}$.

(iii) Our definition of elliptic curve is slightly different from the usual one which requires elliptic curves to be nonsingular. Since we will have nothing to do with singular curves, we would not worry about this difference.

(iv) In characteristic 2 or 3, an elliptic curve should be defined by a more complicated equation to correspond to the standard definition (see [6], page 325). In fact, in characteristic 2, our definition of "singular" is not the right one. Consider a somewhat more general curve defined by an equation $F(X, Y) = 0$ where F is some irreducible polynomial. The usual definition of *singular* is that the curve is singular if there is a point on the curve (*i.e.*, satisfying the above equation) at which both partial derivatives $\partial F/\partial x$ and $\partial F/\partial y$ are zero. In the case of elliptic curves it is easily seen that if the characteristic is not 2, this is equivalent to our definition. This is not the case in characteristic 2, and, in fact, it is easy to see that if we use the definition in terms of partial derivatives, all curves of the form $Y^2 = X^3 + AX + B$ are singular.

In characteristic 3, we will not be considering the most general elliptic curve, but for the class of curves defined by (1) the proofs all work up to Theorem 1.8.7. We will say no more here about these characteristics.

In these notes, **we will consider only non-singular elliptic curves.**

1.3 Polynomial and Rational Functions

We would like to think of the elements of $K[X, Y]$ as defining polynomial functions on E, but clearly two elements of $K[X, Y]$ that differ by a multiple of $Y^2 - X^3 - AX - B$ will define the same function on E. If we wanted to be fancy, we could define polynomials on E to be elements of the quotient ring

$$K[X,Y]/(Y^2 - X^3 - AX - B),$$

where $(Y^2 - X^3 - AX - B)$ is the ideal generated by $Y^2 - X^3 - AX - B$. Of course, the idea of these notes is *not* to be fancy, so we will adopt a more concrete definition of polynomials on E. The point is that $Y^2 = X^3 + AX + B$ on E, so any time we see a power of Y higher than one, we can use this relation to replace it by an expression in X times a power of Y no greater than one. Notice that if we consider polynomials in the functions x and y, the relation $y^2 = x^3 + Ax + B$ holds automatically. We therefore consider polynomials to be elements of $K[x, y]$, the ring of polynomials in the functions x and y.

Definition 1.3.1. A *polynomial on E* is an element of $K[x, y]$. We sometimes denote the ring of polynomials on E by $K[E]$.

An important consequence of this definition is that any polynomial f on E can be written $f(x, y) = v(x) + yw(x)$ for two polynomials v and w of one variable. Also note that using this definition, polynomials are automatically functions on E since x and y are.

Definition 1.3.2. If $f(x, y) = v(x) + yw(x)$ is a polynomial on E, its *conjugate* $\overline{f}$ is the polynomial $\overline{f}(x, y) = v(x) - yw(x)$ and its *norm* is the polynomial $N(f) = f\overline{f}$.

Remark. Without getting too bogged down in foundational considerations, we would like to examine the notion of polynomial a little more carefully. First notice that

$$N(f)(x, y) = v(x)^2 - s(x)w(x)^2 \ ,$$

where $s(x) = x^3 + Ax + B$, so we can think of $N(f)$ as being a function of only one variable, *i.e.*, a member of $K[x]$. In addition, $N(fg) = N(f)N(g)$. Many of the proofs in this part depend on thinking of $N(f)$ in this way because we know a lot about polynomials of only one variable. For example, we might like to know that the representation of a polynomial as $v(x) + yw(x)$ is unique. Suppose $f(x, y) = v(x)+yw(x)$ were the zero function. Then $N(f)$ would have to be the zero function (since the degree of s is odd, and the degrees of v^2 and w^2 are even, the polynomial w would have to be zero and hence so would the polynomial v). From this we easily see that the representation $v(x) + yw(x)$ is unique. In a similar fashion, we can see that $K[x, y]$ has no zero divisors, and that the usual rules used with polynomials hold here.

Now we would like to define a rational function to be the quotient of two polynomials, but we must exercise some care.

Definition 1.3.3. A *rational function on E* is an equivalence class of formal quotients of polynomials f/g (with g not identically zero), where we identify f/g with h/k if $fk = gh$ as polynomials on E. It is easily seen that the set of rational functions on E is a field, which we denote by $K(E)$.

The way to see that $fk = gh$ "as polynomials on E" is to write both fk and gh in the canonical form $v(x)+yw(x)$ using the relation $y^2 = x^3+Ax+B$, and then see if they are equal. While polynomials have values at every finite point of E, rational functions may not have values at all finite points and may have a value at $\mathcal{O}$. Notice that if $r = f/g$ is a rational function, then by multiplying by $\overline{g}/\overline{g}$, we can write $r(x, y) = a(x)+yb(x)$, where now a and b are rational functions of x alone.

Definition 1.3.4. If r is a rational function on E and P is a finite point in E, we say r is *finite at* P if there exists a representation $r = f/g$ where f and g are polynomials on E and $g(P) \neq 0$. If r is finite at P, we put $r(P) = f(P)/g(P)$, and it is trivial to see that this is well-defined.

Exercise 1.3.5. Show that the rational functions that are finite at P form a ring (*i.e.*, sums and products of finite polynomials are finite) and, if you know what it is, show this ring is *local*.

It is somewhat more complicated to define the value of a rational function at $\mathcal{O}$, even if it has one there. The usual way (in calculus) to find the value (or limit) of a rational function at infinity is to compare the degrees of the numerator and denominator. In our case the situation is complicated by the existence of two variables, x and y. While it might seem natural to assign degree 1 to x and y, this would not be consistent with our fundamental relation $y^2 = x^3 + Ax + B$. This relation suggests that the degree of y should be $3/2$ the degree of x. Since we do not want to deal with fractional degrees, we will assign degree 3 to y and degree 2 to x. To avoid confusion, we denote the usual degree of a polynomial f in x alone by $\deg_x(f)$.

Definition 1.3.6. Let $f(x, y) = v(x) + yw(x)$ be a nonzero polynomial on E. Define the *degree of* f by

$$\deg(f) = \max[2 \cdot \deg_x(v), 3 + 2 \cdot \deg_x(w)]. \tag{1.2}$$

If f happens to be a function of only the variable x then its degree as a function on E is twice its usual degree as a function of x, *i.e.*, with the degree of x equal 1.

Lemma 1.3.7. *If f is a polynomial on E then*

$$\deg(f) = \deg_x(N(f)) \ .$$

Proof. If we write $f(x, y) = v(x) + yw(x)$, then $N(f)(x) = v^2(x) - s(x)w^2(x)$, where $s(x) = x^3 + Ax + B$. The lemma then follows from the definition of degree. □

In order to see that this is a useful notion of degree, we must check the fundamental property we expect of degrees.

Proposition 1.3.8. *If f and g are polynomials on E, then*

$$\deg(f \cdot g) = \deg(f) + \deg(g).$$

Proof. Using the lemma, we get

$$\begin{aligned} \deg(fg) &= \deg_x(N(fg)) = \deg_x(N(f)N(g)) \\ &= \deg_x(N(f)) + \deg_x(N(g)) = \deg(f) + \deg(g) \ , \end{aligned}$$

since we certainly know the result for $\deg_x$. □

Note that while we cannot talk about the degree of the numerator (or denominator) of a rational function, the difference between the degree of the numerator and the degree of the denominator is well-defined, *i.e.*, if $r = f/g$, r may also equal h/k and $\deg(f)$ may not equal $\deg(h)$. By the above proposition, however, $\deg(f) - \deg(g) = \deg(h) - \deg(k)$ since $fk = gh$. Therefore we can make the following definitions:

Definition 1.3.9. Suppose $r = f/g$ is a rational function on E. If $\deg(f) < \deg(g)$, we set $r(\mathcal{O}) = 0$. If $\deg(f) > \deg(g)$, we say that

r is not finite at $\mathcal{O}$. If $\deg(f) = \deg(g)$, we must distinguish two cases. If $\deg(f)$ is even, then writing f and g in canonical form, they will have as leading terms (terms of highest degree) ax^d and bx^d respectively (for some $a, b \in K$ and integer d). Then we put $r(\mathcal{O}) = a/b$. Similarly if $\deg(f)$ is odd, the leading terms have the form ayx^d and byx^d, and again we put $r(\mathcal{O}) = a/b$.

Remark. It might seem natural to define the *degree* of a rational function $r = f/g$ to be $\deg(f) - \deg(g)$. If we did this, then r would be finite or infinite at $\mathcal{O}$ depending on whether it had negative or positive degree. The disadvantage of this definition is that it disagrees with the usual one used in algebraic geometry. We will avoid this problem by not defining the degree of a rational function at all.

Exercise 1.3.10. Show that if r and s are rational functions with $r(\mathcal{O})$ and $s(\mathcal{O})$ finite, then $rs(\mathcal{O}) = r(\mathcal{O})s(\mathcal{O})$, and $(r + s)(\mathcal{O}) = r(\mathcal{O}) + s(\mathcal{O})$.

If r is a rational function on E that is not finite at some $P \in E$, we write $r(P) = \infty$ to indicate this.

1.4 Zeros and Poles

From the material of the last section it is easy to define what a zero or pole of a rational function should be.

Definition 1.4.1. Let r be a rational function on E. We say that r has a *zero* at $P \in E$ if $r(P) = 0$ and that r has a *pole* at P if $r(P) = \infty$.

What is not so easy to do is to define the multiplicity of a zero or pole.

Example 1.4.2. Suppose E is given by the equation $Y^2 = X^3 + X$. Then the point $P = (0, 0)$ is in E. Since $x = y^2 - x^3$, it appears that the function x ought to have a zero at P whose multiplicity is twice that of the zero of y at P.

Before we prove a theorem that will show us how to define multiplicities, we want to point out three points on our elliptic curve that will cause us no end of difficulty, beginning here. Remember that we have assumed that E was nonsingular, which means that the polynomial $X^3 + AX + B$ has three distinct roots. Let's call them ω_1, ω_2 and ω_3, and use ω to indicate an arbitrary one. Then E contains three points whose y-coordinate is 0, namely $(\omega_1, 0)$, $(\omega_2, 0)$ and $(\omega_3, 0)$. These three points are called the *points of order two* for reasons that will become apparent in Section 6.

Theorem 1.4.3. *For each point $P \in E$, there is a rational function u, zero at P, with the following property: if r is any rational function not identically zero, then*

$$r = u^d s \tag{1.3}$$

for some integer d and some rational function s that is finite and nonzero at P. Furthermore, the number d does not depend on the choice of the function u.

Proof. There are three cases. First we do the generic case, *i.e.*, we assume that P is not of order 2 and that P is not $\mathcal{O}$. For $P = (a, b)$, we will show we can take $u(x, y) = x - a$. Suppose r has a zero at P.

Then $r = f/g$ with $f(P) = 0$ and $g(P) \neq 0$. If we can decompose $f = u^d s$ as in the above equation, then we can simply divide by g and get the corresponding result for r.

We write $f(x, y) = v(x) + yw(x)$ so $\overline{f}(x, y) = v(x) - yw(x)$. If $\overline{f}(P) = 0$, then since the characteristic is not two and $y(P) = b \neq 0$, we can solve the linear equations

$$\begin{aligned} v(a) + bw(a) &= 0 \\ v(a) - bw(a) &= 0, \end{aligned}$$

to conclude that $v(a) = w(a) = 0$. Since v and w are polynomials in one variable, we get

$$f(x, y) = (x - a) \cdot s_1(x, y)$$

for some polynomial s_1.

If $\overline{f}(P) \neq 0$, then we can multiply f by $\overline{f}/\overline{f}$ to get

$$f(x, y) = \frac{v^2(x) - s(x)w^2(x)}{\overline{f}(x, y)}$$

where $s(x) = x^3 + Ax + B$. Now $f(P) = 0$ and $\overline{f}(P) \neq 0$ implies

$$v^2(x) - s(x) \cdot w^2(x) = 0 \text{ for } x = a,$$

and the polynomial on the left is a polynomial in one variable. Again we conclude that

$$f(x, y) = (x - a) \cdot s_1(x, y) \ ,$$

where this time s_1 is some rational function that is finite at P. In either case, if $s_1(P) = 0$, we can continue the process. To see that it eventually comes to an end, note that if $f(x,y) = (x-a)^d s_1(x,y)$, then $N(f)(x) = (x-a)^{2d} N(s_1)(x)$. We know that $N(s_1)(x)$ does not have a pole at a so we can see that $2d$ must be less than the degree of $N(f)$ as a function of x alone.

Thus if r has a zero at $P=(a,b)$, we can take $u(x,y) = x-a$. If r has a pole at P, then $1/r$ has a zero at P, and the same u still works (with d negative). If r has neither a zero nor a pole at P, then we can take $d=0$ and $s=r$, and what u is is immaterial. Thus in the generic case, we can take $u(x,y) = x-a$.

Now we assume that P is a point of order two, say $P=(\omega_1, 0)$. We show that we can take $u(x,y)=y$ in this case. As above, if r has a zero at P, we can assume $r=f/g$ and $f(P)=0$. Now $f(\omega_1,0)=0$ implies $v(\omega_1)=0$ where $f(x,y)=v(x)+yw(x)$. Hence we can write $v(x) = (x-\omega_1)v_1(x)$ for some polynomial v_1. Since the roots of $s(x)$ are distinct, $(x-\omega_2)$ and $(x-\omega_3)$ do not vanish at P, so we get

$$\begin{aligned} f(x,y) &= (x-\omega_1)v_1(x) + yw(x) \\ &= \frac{(x-\omega_1)(x-\omega_2)(x-\omega_3)v_1(x) + yw_1(x)}{(x-\omega_2)(x-\omega_3)} \\ &= \frac{y^2 v_1(x) + yw_1(x)}{(x-\omega_2)(x-\omega_3)} \\ &= y\left[\frac{yv_1(x) + w_1(x)}{(x-\omega_2)(x-\omega_3)}\right] \quad , \end{aligned}$$

where $w_1(x) = (x-\omega_2)(x-\omega_3)w(x)$. Now if the function in brackets still vanishes at P, we can do the process over again to the polynomial $w_1(x) + yv_1(x)$. This process must also terminate since in every other step we factor $x-\omega_1$ from v, which can contain only finitely many such factors. Hence in the case of points of order two, we can take $u(x, y) = y$.

Finally in the case $P = \mathcal{O}$, we show that $u(x, y) = x/y$ works. Suppose $r = f/g$ and $r(\mathcal{O}) = 0$. This means that $\deg(f) - \deg(g) = d < 0$. Since $\deg(y) - \deg(x) = 1$, $\deg(y^d f) = \deg(x^d g)$, and $(y/x)^d r$ will be finite and nonzero at the identity. Since

$$r = (x/y)^d \left[(y/x)^d r \right] ,$$

we see that we can take $u(x, y) = x/y$ at the identity.

To see that the number d is unique, suppose that u and u' are both rational functions satisfying the conditions of the theorem. This means that we can write $u = (u')^e s$ and $u' = u^f t$, so $u = u^{ef}(t^e s)$. If $ef \neq 1$, then by dividing this equation by u and plugging in P, we get $1 = 0$. We therefore must have $e = f = 1$. Thus if r is any rational function not identically zero that vanishes at P, we can write $r = u^d s = (u')^d t$. □

The above theorem allows us to make the following definitions:

Definition 1.4.4. A function u that satisfies the above theorem is called a *uniformizing variable* or *uniformizer* at P. If r is a rational function and $r = u^d s$, where u is a uniformizing variable at P, we say that the *order* of r at P is d and write

$$\mathrm{ord}_P(r) = d .$$

We define the *multiplicity* of a zero to be the order of the function and the *multiplicity* of a pole to be the negative of the order. If a zero or pole has multiplicity one, two, or three we say it is *simple*, *double*, or *triple*, respectively.

Lemma 1.4.5. *For any $P \in E$, $r_1, r_2 \in K(E)$,* ord_P *satisfies*

$$\operatorname{ord}_P(r_1 r_2) = \operatorname{ord}_P(r_1) + \operatorname{ord}_p(r_2) \ .$$

Proof. If $r_1 = u^{d_1} s_1$ and $r_2 = u^{d_2} s_2$ with s_1, s_2 finite and nonzero at P then $r_1 r_2 = u^{d_1+d_2} s_1 s_2$ and $s_1 s_2$ finite and nozero at P. □

Example 1.4.6.

(i) Let $P \in E$ and suppose $P = (k, l)$ with $k, l \in K$ and $l \neq 0$. Let $u = x - k$. Since u is a uniformizer at P, we see $\operatorname{ord}_P(u) = 1$. Now $P' = (k, -l)$ is also a point of E, and clearly $\operatorname{ord}_{P'}(u) = 1$. It is also clear that u has order zero at every other finite point. We see that u has a pole at $\mathcal{O}$, and since $\deg(u) = 2$, we get that $\operatorname{ord}_{\mathcal{O}}(u) = -2$. Summing up, we see that u has two simple zeros, and a single double pole.

(ii) Now consider the function y. We have seen that y is a uniformizing variable at the three points $(\omega_1, 0)$,$(\omega_2, 0)$, and $(\omega_3, 0)$, so it has a zero of multiplicity one at these three points. Also y has order zero at every other point except $\mathcal{O}$. Since y has degree three, it has a pole of multiplicity three at $\mathcal{O}$.

(iii) Let $P_i = (\omega_i, 0)$ and $f_i = x - \omega_i$. Then $f_1 f_2 f_3 = y^2$ and $\operatorname{ord}_{P_i}(f_i) = 2$.

(iv) Finally take $u = x/y$. We leave it to the "interested reader" to show that u has a zero of multiplicity one at $\mathcal{O}$, zeros also of multiplicity one at the two points $(0, \sqrt{B})$ and $(0, -\sqrt{B})$ and simple poles at the three points of order two (if $B = 0$, there is a simple zero at $\mathcal{O}$, a simple zero of multiplicity one at $(0, 0)$, and poles of multiplicity one at the points $(\sqrt{-A}, 0)$ and $(-\sqrt{-A}, 0)$).

These examples suggest the following theorem, which is a sort of baby Riemann-Roch Theorem in that it places restrictions on what kind of zeros and poles a rational function can have:

Theorem 1.4.7. *Let r be a rational function on E. Then*

$$\sum_{P\in E} \operatorname{ord}_P(r) = 0 \ .$$

Before we give its proof, we prove some lemmas, which are of interest themselves.

Lemma 1.4.8. *Let $f \in k(E)$, and $P = (a, b)$. Assume f can be written $f = (x-a)^d u$ in $K(x, y)$ with $u(P)$ finite and nonzero. Then if $b \neq 0$, $d = \operatorname{ord}_P(f)$; if $b = 0$, $\operatorname{ord}_P(f) = 2d$.*

Proof. $f = (x-a)^d u$ holds in $K(E)$. the conclusion comes from the order of $(x-a)$ at P. □

The observation in the above lemma is most useful for $f \in k(X)$.

Lemma 1.4.9. *Let f be a polynomial on E. The sum of the multiplicities of the zeros of f equals the degree of f.*

Proof. Let $\deg(f) = n$. By Lemma 1.3.7, $\deg_x(N(f)) = n$, so we can write

$$N(f)(x) = f\overline{f}(x) = (x-a_1)(x-a_2)\cdots(x-a_n),$$

where the a_i are elements of K that may not be distinct. If $a_i \neq \omega$, then $(x - a_i)$ has two distinct roots on E. If $a_i = \omega$, then $x - a_i$ has only one root on E, but it has multiplicity two. Thus we can conclude that $f\overline{f}$ has precisely $2n$ roots on E, counting multiplicities. But clearly f and $\overline{f}$ have the same number of roots on E, so the sum of the multiplicities of the zeros of each must be n. □

Now we give the proof of Theorem 1.4.7.

Proof. It suffices to prove the result for a polynomial f. We know that

$$\sum_{P \in E - \mathcal{O}} \operatorname{ord}_p(f)$$

is the sum of the multiplicities of the zeros of f. On the other hand, by definition, $\operatorname{ord}_{\mathcal{O}}(f)$ is the negative of the degree of f, so the theorem follows from the lemma. □

We will need the next two lemmas in several places.

Lemma 1.4.10. *Let f be a nonconstant polynomial on E. Then f must have at least two simple zeros or one double one at finite points of E.*

Proof. If f is nonconstant, then it must involve an x or a y. Since $\deg(x) = 2$ and $\deg(y) = 3$, this result follows from the previous lemma. □

The following exercise is in much the same spirit as the above lemma:

Exercise 1.4.11. Since K is algebraically closed, E must have an infinite number of points. Show that if two rational functions agree on an infinite number of points of E, then they are equal.

Lemma 1.4.12. *A rational function without finite poles is a polynomial.*

Proof. We write $r = a + yb$ where a and b are rational functions of x. If r has no finite poles, then clearly $\overline{r} = a - yb$ has no finite poles. Hence $r + \overline{r} = 2a$ has no finite poles. If a had a pole as a function of x, then it would have one as a function on E. Thus a is a polynomial. This implies that $yb = r - a$ has no finite poles. Hence $(yb)^2 = sb^2$ has no finite poles where $s(x) = x^3 + Ax + B$. If b has a pole, then it can be written $b = f/g$ where $g(x) = 0$ for some $x \in K$.

In this case, $b^2 = f^2/g^2$ and g^2 has a double root at some x. The only way for sb^2 not to have a pole at P is for s to have a double zero at x, and since E is nonsingular, this is not the case. Therefore b has no finite poles. Finally we see that b must also be a polynomial, so r is a polynomial. □

There is an extension of the idea of rational function that we will need later.

Definition 1.4.13. A *rational map* F on E is a pair (r, s) where r and s are rational functions on E such that

$$s^2 = r^3 + Ar + B \ .$$

If we make the convention that $F(P) = \mathcal{O}$ if r and s are not finite at P, we see that F actually defines a map from E to E by $F(P) = (r(P), s(P))$ since r and s must have poles at the same points.

Remark. There is an amusing way of looking at rational maps that will actually be useful. Given the field K, we form the elliptic curve E using the equation

$$Y^2 = X^3 + AX + B. \tag{1.4}$$

Now suppose we consider the field of rational functions $K(E)$. Then we can use the same equation to form a new elliptic curve, which we might denote by $E(K(E))$. Now $K(E)$ may not be algebraically closed, and by our convention, the points of $E(K(E))$ have coordinates in the algebraic closure of $K(E)$. The finite points whose coordinates lie in $K(E)$ (*i.e.*, the $K(E)$-rational points) are precisely the rational maps. We can think of the identity of this curve, call it $\mathcal{O}_M$, as the "map" with the constant value $\mathcal{O}$.

1.5 Divisors and Lines

It is not hard to imagine that it would be convenient to have a device to keep track of the zeros and poles of a rational functional r. One idea is to use lists

$$[(P_1, m_1), (P_2, m_2), \ldots, (P_n, m_n)]$$

where r has order m_i at P_i. It turns out to be better to consider a formal sum

$$m_1 \langle P_1 \rangle + m_2 \langle P_2 \rangle + \cdots + m_n \langle P_n \rangle = \sum_{i=1}^{n} m_i \langle P_i \rangle \ .$$

The right way to do this is to use the notion of a free Abelian group generated by a given set. We recall the definition here.

Definition 1.5.1. Let S be any set. *The free Abelian group generated by* S is the set of finite formal linear combinations

$$\sum_{s \in S} m(s)s,$$

where $m(s) \in \mathbb{Z}$, and $m(s) = 0$ except for finitely many $s \in S$. The addition is also formal; simply juxtapose and collect terms. For example,

$$(m_1 s_1 + m_2 s_2) + (n_1 s_1 + n_3 s_3) = (m_1 + n_1)s_1 + m_2 s_2 + n_3 s_3 \ .$$

Definition 1.5.2. Let E be an elliptic curve over an algebraically closed field K. The *group of divisors of* E is the free Abelian group generated by the points of E. We denote it by Div(E). To distinguish the point P from the divisor whose sole nontrivial entry is P with coefficient 1, we denote this divisor by $\langle P \rangle$. If $\Delta = \sum_{P \in E} m(P) \langle P \rangle$ is a divisor, then we define its *degree* by

$$\deg(\Delta) = \sum_{P \in E} m(P) \in \mathbb{Z} \ .$$

If r is a nonzero rational function on E, we associate a divisor to r by the following equation:

$$\operatorname{div}(r) = \sum_{P \in E} \operatorname{ord}_p(r) \langle P \rangle \ \ .$$

Remarks.

(i) We should observe that a rational function has a finite number of zeros and poles. This can be seen from Lemma 1.4.9.

(ii) If two rational functions have the same divisor, then Lemma 1.4.12 implies that their quotient is constant. Thus one way to prove that two functions are equal is to show that they have the same divisor, then to show they agree at any one point of E. Usually the only point on E that we can get our hands on is $\mathcal{O}$, and frequently functions have poles at $\mathcal{O}$. In this case, we can compare leading coefficients, defined below.

Definition 1.5.3. Let r be a rational function and suppose

$$\operatorname{ord}_{\mathcal{O}}(r) = d.$$

Then we define the *leading coefficient* of r to be

$$[(x/y)^d \cdot r](\mathcal{O}) \ .$$

Exercise 1.5.4. Show that if two rational functions have the same divisor and the same leading coefficient, they are equal.

Example 1.5.5.

(i) Let $P = (a, b) \in E$ with $b \neq 0$, and $r_1 = (x - a)$. Then we have seen that r_1 has simple zeros at P and $P' = (a, -b)$ and a pole of multiplicity two at $\mathcal{O}$. Therefore $\operatorname{div}(r_1) = \langle P \rangle + \langle P' \rangle - 2\langle \mathcal{O} \rangle$ where we have used $-2\langle \mathcal{O} \rangle$ for $+(-2)\langle \mathcal{O} \rangle$.

(ii) Let $r_2 = y$. If we let $P_i (i = 1, 2, 3)$ be the points of order two, then

$$\operatorname{div}(r_2) = \langle P_1 \rangle + \langle P_2 \rangle + \langle P_3 \rangle - 3\langle \mathcal{O} \rangle \ .$$

(iii) We take $r_3 = x/y$ and $Q = (0, \sqrt{B})$ and $Q' = (0, -\sqrt{B})$, so

$$\operatorname{div}(r_3) = \langle Q \rangle + \langle Q' \rangle - \langle P_1 \rangle - \langle P_2 \rangle - \langle P_3 \rangle + \langle \mathcal{O} \rangle \ .$$

Definition 1.5.6. We say a divisor Δ is *principal* if $\Delta = \operatorname{div}(r)$ for some rational function r. If $\Delta_1 - \Delta_2$ is principal, we say Δ_1 and Δ_2 are *linearly equivalent* or *in the same divisor class*, and write $\Delta_1 \sim \Delta_2$.

Proposition 1.5.7. *If r_1 and r_2 are rational functions on E, then $\operatorname{div}(r_1 r_2) = \operatorname{div}(r_1) + \operatorname{div}(r_2)$.*

Proof. Direct consequence of lemma 1.4.5. □

Definition 1.5.8. By the above proposition, the set of principal divisors forms a subgroup of $\operatorname{div}(E)$, which we denote by $\operatorname{Prin}(E)$. We also define $\operatorname{div}^0(E)$ to be the subgroup of divisors of degree 0. (It is trivial to see it is a subgroup.)

One of our goals in this section is to study which divisors are principal, *i.e.*, what zeros and poles a rational function can have. This is equivalent to studying the divisors that are *not* principal. These divisors are represented by the elements of the group

$$\operatorname{Pic}(E) = \operatorname{div}(E) / \operatorname{Prin}(E)$$

$\operatorname{Pic}(E)$ is called the *Picard* or *divisor class group of E*. Actually we can study a smaller group to investigate which divisors are principal. Theorem 1.4.7 implies $\operatorname{Prin}(E) \subseteq \operatorname{div}^0(E)$, so we may as well look at the divisors of degree zero that are not principal, *i.e.*, the group

$$\operatorname{Pic}^0(E) = \operatorname{div}^0(E) / \operatorname{Prin}(E) \ .$$

$\operatorname{Pic}^0(E)$ is called the *degree zero part of the Picard* (or *divisor class*) *group of E*. We are going to show that $\operatorname{Pic}^0(E)$ is in one-to-one correspondence with the points of E. We need some more definitions first.

Definition 1.5.9. If $\Delta = \sum\limits_{P \in E} m(P) \langle P \rangle$ is a divisor, we define its *norm* by

$$|\Delta| = \sum_{P \in E - \mathcal{O}} |m(P)| \ .$$

For example, a divisor of norm one looks like $\pm \langle P \rangle + n \langle \mathcal{O} \rangle$ where n is some arbitrary integer. Also if Δ is the divisor of some polynomial f, then $|\Delta|$ is the sum of the multiplicities of the zeros of f, which is the degree of f.

Definition 1.5.10. A *line* on E is a polynomial of the form

$$\ell(x, y) = \alpha x \ + \beta y + \gamma,$$

for some $\alpha, \beta, \gamma \in K$ with not both α and β zero.

If a point P is a zero of the line ℓ, we say ℓ is a line *through* P, and P is *on* ℓ.

The main result on lines is the following:

Lemma 1.5.11. *If ℓ is a line with divisor Δ, then $|\Delta| = 2$ or 3.*

Proof. ℓ is a polynomial of degree 2 (if $\beta = 0$) or 3 (if $\beta \neq 0$). Hence by Lemma 1.4.9, the sum of the multiplicities of the zeros of ℓ is 2 or 3, and this sum is precisely $|\Delta|$. □

Exercise 1.5.12. Show that the possible divisors of a line are the following, where P, Q, and R are distinct:

(i) $\operatorname{div}(P) = \langle P \rangle + \langle Q \rangle + \langle R \rangle - 3 \langle \mathcal{O} \rangle$.

(ii) $\operatorname{div}(P) = 2 \langle P \rangle + \langle Q \rangle - 3 \langle \mathcal{O} \rangle$.

(iii) $\operatorname{div}(P) = 3\langle P\rangle - 3\langle \mathcal{O}\rangle$.

(iv) $\operatorname{div}(P) = \langle P\rangle + \langle Q\rangle - 2\langle \mathcal{O}\rangle$.

(v) $\operatorname{div}(P) = 2\langle P\rangle - 2\langle \mathcal{O}\rangle$.

Show all of these cases actually occur. (Hint: In part (iii), P is an inflection point of the curve.)

Proof. Since a line has degree 2 or 3 at $\mathcal{O}$, it is clear that a divisor of a line is one of the above. The first three cases correspond to a line with some terms in y, while the two last cases correspond to a line equation in $K(x)$.

(i) example: $y = 0$.

(ii) example: $y = a$, with a such that $x^3 - Ax - B - a^2$ has a double root (which is in general possible as there is one constraint to satisfy).

(iii) $y = ax + b$, with a, b such that $x^3 - Ax - B - (ax+b)^2$ has a triple root

(iv) $x - a$ with a not a root of $x^3 - Ax - B$.

(v) $x - a$ with a a root of $x^3 - Ax - B$.

□

The next theorem is the main result of this section. It has an amazingly simple proof. If $P = (a, b)$, we use the notation $-P$ for $(a, -b)$. The reason for this will become clear in the next section. Observe that the divisor $\langle P\rangle + \langle -P\rangle - 2\langle \mathcal{O}\rangle$ is always principal (if P is of order two then P and $-P$ may not be distinct).

Theorem 1.5.13. *(Linear Reduction) Let $\Delta \in \operatorname{div}(E)$. Then there is $\tilde{\Delta} \in \operatorname{div}(E)$ with $\tilde{\Delta} \sim \Delta$, $\deg(\tilde{\Delta}) = \deg(\Delta)$ and $|\tilde{\Delta}| \leq 1$.*

Proof. Suppose $\Delta = \sum_{P \in E} n(P) \langle P \rangle$, and Q and R appear in Δ with nonzero coefficients of the same sign. Let ℓ be the line through Q and R. Then depending on the sign of the coefficient of Q (or R), $\Delta + \operatorname{div}(P)$ or $\Delta - \operatorname{div}(P)$ will have $|n(Q)|$ and $|n(R)|$ reduced by one if ℓ has three distinct zeros. By the lemma, we will have at worst increased the coefficient of one other point by one. If ℓ has only two distinct zeros, then we will have decreased $|n(Q)|$ or $|n(R)|$ by one, and increased no other coefficient at all. So we will have produced a divisor, say Δ_1, with $\Delta_1 \sim \Delta, \deg(\Delta_1) = \deg(\Delta)$ and $|\Delta_1| < |\Delta|$.

After doing this linear reduction a finite number of times, we get a divisor Δ' linearly equivalent to Δ, of the same degree as Δ, and

$$\Delta' = n_1 \langle P \rangle - n_2 \langle Q \rangle + n \langle \mathcal{O} \rangle \,,$$

where n_1 and n_2 are nonnegative integers and n is an integer we do not care about.

Suppose $n_1 > 1$. Consider the line

$$\ell(x, y) = m(x - a) - (y - b)$$

with $P = (a, b)$. If P is not of order two, $b \neq 0$ and $\operatorname{ord}_P(\ell) = \operatorname{ord}_P(f)$ where $f = \ell\,\overline{\ell} = [m(x - a) + b]^2 - x^3 - Ax - B$.

By lemma 1.4.8, $\operatorname{ord}_P(f)$ can be computed as in $K(x)$. Since $f'(a) = 2\,b\,m - 3a^2 - A$, assuming $b \neq 0$, P has multiplicity two if

$$m = \frac{3a^2 + A}{2\,b} \,.$$

This line has divisor $2\langle P\rangle + \langle S\rangle - 3\langle \mathcal{O}\rangle$. By subtracting it, we can reduce n_1 and $|\Delta'|$.

If $P = (0, \omega)$ is of order two, the line $\ell = x - \omega$ has divisor $2\langle P\rangle - 2\langle \mathcal{O}\rangle$ and can be subtracted to reduce n_1.

(Another way to do this computation is to use the derivation of Definition 1.8.1.)

We can similarly reduce n_2. Eventually we are done, or we arrive at

$$\langle P\rangle - \langle Q\rangle + ?\langle \mathcal{O}\rangle \quad .$$

The line $\ell = x - a$ has divisor $\langle P\rangle + \langle R\rangle - 2\langle \mathcal{O}\rangle$ or $2\langle P\rangle - 2\langle \mathcal{O}\rangle$. Thus by subtracting it, we are reduced to a previous case. □

The next two corollaries and Proposition 1.6.7 are analogs of the Riemann-Roch Theorem. They give a rather precise description of which divisors are principal.

Corollary 1.5.14. *For each $\Delta \in Div^0(E)$, there is a unique point $P \in E$ such that*

$$\Delta \sim \langle P\rangle - \langle \mathcal{O}\rangle \quad .$$

Proof. The theorem tells us that Δ is equivalent to a divisor of norm 1, *i.e.*, a divisor $\pm\langle P\rangle + n\langle \mathcal{O}\rangle$. If the sign of $\langle P\rangle$ is not plus, by subtracting the line with divisor $\langle P\rangle + \langle Q\rangle - 2\langle \mathcal{O}\rangle$, we can change the sign. Since we are given that degree of Δ is zero, the coefficient of $\mathcal{O}$ must be -1, so the only thing to check is whether P is unique. Suppose $\Delta \sim \langle Q\rangle - \langle \mathcal{O}\rangle$ also. Then $\langle Q\rangle \sim \Delta + \langle \mathcal{O}\rangle \sim \langle P\rangle$, so there would have to be a rational function r with $\mathrm{div}(r) = \langle P\rangle - \langle Q\rangle$. By theorem 1.5.13, we can see that if this were the case, there would have to be a rational function r whose divisor is $\langle S\rangle - \langle \mathcal{O}\rangle$ for some $S \in E$. Clearly r would have no finite poles so by Lemma 1.4.12, r

would have to be a polynomial. But then r would be a polynomial with a single finite zero, which by Lemma 1.4.10 is impossible, so we must have $\langle P\rangle = \langle Q\rangle$. □

Define a map $\overline{\sigma}\colon \operatorname{div}^0(E) \to E$ by $\overline{\sigma}(\Delta) = P$ where P is the unique point with $\Delta \sim \langle P\rangle - \langle \mathcal{O}\rangle$. Since $\forall\, r \in K(E)$, $\operatorname{div}(r) \sim 0$, $\overline{\sigma}(\operatorname{div}(r)) = \mathcal{O}$, and we see that $\overline{\sigma}$ induces a map

$$\sigma : \operatorname{Pic}^0 \to E\ .$$

Corollary 1.5.15. *σ is a bijection.*

Proof. σ is surjective since $\sigma(\langle P\rangle - \langle \mathcal{O}\rangle) = P$. For $d_1, d_2 \in \operatorname{div}^0(E)$, $\overline{\sigma}(d_1) = \overline{\sigma}(d_2) = P$ means $d_1 \sim \langle P\rangle - \langle \mathcal{O}\rangle \sim d_2$, hence $d_1 = d_2$ in Pic^0, hence σ is injective. □

1.6 The Group Laws

We are going to give the group addition laws for an elliptic curve in their algebraic formulation. This is in keeping with our policy of being explicit and computational. It does, however, suffer from two disadvantages. One is that it is somewhat unmotivated, and the other is that the "(direct) verification (of the associativity law) is a pain" ([3], page 40). Another approach is geometric (see [6], page 55) and is a little more motivated, but associativity is still difficult (although it can be done, see [1], page 125). We will use this approach to motivate the algebraic equations. In his book, Lang gives a beautiful definition using doubly periodic functions, but this method only works when K is the field of complex numbers. Finally there is a way to define addition using divisors that makes associativity trivial. We are ultimately going to use this approach by showing it is equivalent to the algebraic formulas presented here.

Remark. The basic idea behind addition on an elliptic curve is that a line will intersect the curve no more than three times. We describe roughly how this works; the formal definitions follow. First we make $\mathcal{O}$ act as the zero or identity element of the group. Then we make $(a, -b)$ be the negative of (a, b). Finally, if the points P, Q, and R are on a line, we define the addition so that $P + Q + R = \mathcal{O}$.

First of all, suppose $P \neq Q$ or $-Q$ and let ℓ be the line through P and Q and let R be the third zero of ℓ, which is easily seen to be finite. Write $P = (a, b)$ and $Q = (c, d)$ so ℓ can be written

$$\ell = m(x - a) - (y - b) \ ,$$

where $m = (d - b)/(c - a)$. We have seen that since (a, b) is a zero of ℓ and is on the curve, a is a zero of the polynomial

$$f(x) = [m(x - a) + b]^2 - x^3 - Ax - B. \tag{1.5}$$

It is trivial to see that a and c are zeros of f, but what is the third zero of f? Let e be this third zero. Writing $f(x) = (x - a)(x - c)(x - e)$, we see that the coefficient of x^2 in f is $a + c + e$. Using Equation (1.5), we see that m^2 is the coefficient of x^2, so $a + c + e = m^2$ or $e = -a - c + m^2$. To get the y coordinate of R, we just plug this back into the equation of the line; the y coordinate of R is $m(e - a) + b$.

If $P = Q$, then we use the line tangent to P, *i.e.*, the line with a double zero at P. We have seen that this line has the usual equation with $m = (3a^2 + A)/2b$.

Summing up, to add two distinct nonzero points that are not negatives, draw the line through them and then “flip” the third point of intersection about the x axis (*i.e.*, send (a, b) into $(a, -b)$) to get their sum. If the points to be added are the same, use the line tangent to get the point to be flipped.

Now we give the formal definitions. As usual, we let E be a nonsingular elliptic curve over an algebraically closed field K given by the equation $Y^2 = X^3 + AX + B$. We now define the structure of an Abelian group on E.

Definition 1.6.1. We let the identity $\mathcal{O}$ be the zero of the group (which explains its notation). So for any point $P \in E$,

$$P + \mathcal{O} = \mathcal{O} + P = P \ .$$

For $P = (k, l) \in E$, we define $-P$ to be $(k, -l) \in E$, so

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

Now suppose P_1 and P_2 are not $\mathcal{O}$, and $P_1 \neq -P_2$. Let $P_1 = (k_1, l_1)$ and $P_2 = (k_2, l_2)$. If $k_1 \neq k_2$ (so $P_1 \neq P_2$), define

$$\lambda = \frac{l_2 - l_1}{k_2 - k_1},$$

while if $k_1 = k_2$ (so $P_1 = P_2$ since we have assumed $P_1 \neq -P_2$), define

$$\lambda = \frac{3k_1^2 + A}{2l_1} \ .$$

Define $P_1 + P_2 = (k_3, l_3)$ by

$$k_3 = -k_1 - k_2 + \lambda^2$$

and

$$l_3 = -l_1 - \lambda(k_3 - k_1) \ .$$

We will call the addition formula in the case $P_1 \neq P_2, -P_2, \mathcal{O}$ (*i.e.*, $\lambda = (l_2 - l_1)/(k_2 - k_1)$) the *generic* addition formula since it is the one to use for practically all pairs of points.

Remarks.

(i) In some real sense, the generic formula works all of the time. We can, for example, use it in the case $P_1 = P_2$, not a point of order two, if we believe the following computation:

$$\begin{aligned} \lambda &= \frac{l_2 - l_1}{k_2 - k_1} \cdot \frac{l_2 + l_1}{l_2 + l_1} \\ &= \frac{[k_2^3 - k_1^3] + A(k_2 - k_1)}{(k_2 - k_1)(l_2 + l_1)} \\ &= \frac{k_2^2 + k_1 k_2 + k_2^2 + A}{l_1 + l_2}\ , \end{aligned} \tag{1.6}$$

and if $k_2 = k_1 = k$ and $l_2 = l_1 = l \neq 0$, this becomes $\frac{3(k)^2+A}{2l}$. Geometrically, what this means is that the slope of the line through P_1 and P_2 (namely $\frac{l_2-l_1}{k_2-k_1}$) becomes the slope of the line tangent to P_1 (namely $\frac{3(k)^2+A}{2l}$) as P_1 goes to P_2. There are a number of other cases (e.g., $P = \mathcal{O}$, or $P_1 = -P_2$), and it is not too difficult to work them all out. But what do they mean?

We are really showing here the important fact that addition is a rational function. But a rational function on what? Unfortunately, addition is defined on $E \times E$, which is not an elliptic curve but a product of elliptic curves. It would be very convenient for us to know that addition is rational, but that would involve a more general definition of rational function on a more general algebraic geometric object. All this would lead us too far afield, so we have decided on a different approach.

It might appear that addition is rational simply because it is given by rational formulas. The difficulty is that it is given by *different* rational formulas for different cases. If it is to be obvious that a function is rational, it must be given by the same rational expression at every point it is defined or at least by expressions that are "rationally related", *i.e.*, if $r = f/g$, r may also equal h/k if $fk = gh$.

(ii) One might prefer to use the last expression in Equation (1.6) to define λ in the case $P_1 = P_2$. This would have the advantage of working not only when $P_1 = P_2$, but whenever $l_1 \neq -l_2$. It might almost appear that this expression for λ would work whenever $P_1 \neq \mathcal{O}$ or $-P_2$. However, there are two other points on E besides $-P_2$ whose y coordinate is $-l_2$, so we would still have the same number of special cases.

(iii) Notice that the three points of order two, $(\omega_1, 0), (\omega_2, 0)$ and $(\omega_3, 0)$, satisfy $2.P = \mathcal{O}$, and these are the only points (except for $\mathcal{O}$) which do so. This follows since the definition of $-P$ tells us that any point with $P = -P$ must have second coordinate 0, and the three points of order two are the only points that do.

(iv) If P and Q are any points not both $\mathcal{O}$, then we can find a line ℓ whose divisor is

$$\langle P \rangle + \langle Q \rangle + \langle R \rangle - 3\langle \mathcal{O} \rangle \quad ,$$

and then R is $-P-Q$. This is true even if $Q = \pm P$ or $\mathcal{O}$.

(v) Recall the elliptic curve $E(K(E))$ of rational maps of E. Since the field $K(E)$ may not be algebraically closed, it is not immediately clear that the sum of two rational maps is again a rational map; it may be something whose coordinates lie in the algebraic closure of $K(E)$. However, an examination of the algebraic formulas defining addition quickly shows that this is not the case, *i.e.*, the sum of two rational maps *is* again a rational map.

Now we state the basic theorem about addition.

Theorem 1.6.2. *The elliptic curve with the addition defined above forms an Abelian group.*

All of the group axioms are trivial to check except associativity, which will follow from the next proposition. Recall the bijection

$$\sigma : \operatorname{Pic}^0(E) \to E,$$

defined in the previous section. Let $\kappa = \sigma^{-1}$, so $\kappa(P)$ is the linear equivalence class of the divisor $\langle P \rangle - \langle \mathcal{O} \rangle$. Clearly $\kappa(\mathcal{O})$ is the zero linear equivalence class.

Proposition 1.6.3. $\kappa(P+Q) = \kappa(P) + \kappa(Q)$.

Proof. The result is trivial if $P = Q = \mathcal{O}$, so we suppose that P and Q are not both $\mathcal{O}$. Let ℓ be the line through P and Q. As we remarked above (iv), we can write the divisor of ℓ as

$$\operatorname{div}(\ell) = \langle P \rangle + \langle Q \rangle + \langle R \rangle - 3 \langle \mathcal{O} \rangle \ .$$

Let l' be the line through R and $-R$, so

$$\operatorname{div}(\ell') = \langle R\rangle + \langle -R\rangle - 2\langle \mathcal{O}\rangle \ .$$

We have seen that $R = -P - Q$ so $-R = P + Q$ and

$$\operatorname{div}(\ell'/\ell) = \langle P+Q\rangle - \langle P\rangle - \langle Q\rangle + \langle \mathcal{O}\rangle \sim 0$$

since ℓ'/ℓ is a rational function. Rewriting this slightly yields

$$(\langle P+Q\rangle - \langle \mathcal{O}\rangle) - (\langle P\rangle - \langle \mathcal{O}\rangle) - (\langle Q\rangle - \langle \mathcal{O}\rangle) \sim 0,$$

which translates to $\kappa(P+Q) - \kappa(P) - \kappa(Q) = 0$ as desired. □

Corollary 1.6.4. *Addition on an elliptic curve is associative.*

This follows because the addition in $\operatorname{Pic}^0(E)$ is certainly associative.

We have proved that κ is a homomorphism, so clearly σ is a homomorphism. Since $\overline{\sigma}$ is the composition of σ and projection from div^0 to $\operatorname{Pic}^0 = \operatorname{div}^0/\operatorname{Prin}$, it too is a homomorphism. There is a simpler way of looking at $\overline{\sigma}$.

Definition 1.6.5. Define a map *sum* from $\operatorname{div}(E)$ to E by

$$\operatorname{sum}\left(\sum n(P)\langle P\rangle\right) = \sum n(P)P.$$

Exercise 1.6.6. Show that $\overline{\sigma}$ is merely sum restricted to div^0.

We can use all this to prove an extremely useful result.

Proposition 1.6.7. *Let $\Delta = \sum_{P\in E} n(P)\langle P\rangle$ be a divisor. Then Δ is principal if and only if $\deg(\Delta) = \sum_{P\in E} n(P) = 0$ and $\operatorname{sum}(\Delta) = \sum_{P\in E} n(P)P = \mathcal{O}$.*

Proof. We have already proved the part about the degree, so we can assume that $\deg(\Delta) = 0$. Recall that $\overline{\sigma}(\Delta) = P$ where P is the unique point with $\Delta \sim \langle P \rangle - \langle \mathcal{O} \rangle$. Hence $\Delta \sim 0 \Leftrightarrow \overline{\sigma}(\Delta) = \mathcal{O} \Leftrightarrow$ sum $(\Delta) = \mathcal{O}$, which is what we wanted to prove. □

1.7 Multiplication by n

What we really are interested in is the function $[n] : P \mapsto n \cdot P$, the point P added to itself n times. More precisely, we are interested in the two functions $g_n(P) = x(n \cdot P)$ and $h_n(P) = y(n \cdot P)$, *i.e.*, $n \cdot P = (g_n(P), h_n(P))$.

The next theorem is our version of the fact that addition is rational. Recall that the rational maps on E form an elliptic curve themselves. When we write the sum of two rational maps, we mean the sum on this elliptic curve. Suppose we make the convention that the constant map $\mathcal{O}_M$ whose value is $\mathcal{O}$ everywhere is a rational map. Then the following theorem says that the pointwise sum of rational maps is a rational map. This result is used implicitly all the time: for instance, it ensures that it is equivalent to define (g_n, h_n) by $(g_n, h_n) = (x, y) + (x, y) + \ldots + (x, y)$ (n times) or by $(g_n, h_n)(P) = nP = P + \ldots + P$ (n times). This second definition may not even make sense as a rational map definition, since the expression used to define the various sums that compose it depends on whether the operands of the sum are equal or not, which in turn depends on the value of P.

Theorem 1.7.1. *Let F and G be rational maps on E. If $K = F + G$, then $K(P) = F(P) + G(P)$.*

Proof. The point of the theorem is that even if F is not $G, -G$, or $\mathcal{O}_M$, $F(P)$ may equal $G(P), -G(P)$, or $\mathcal{O}$, so there are some things to be checked.

If any of F, G, or $F+G$ is $\mathcal{O}_M$, the result is trivial, so we will exclude these trivial cases.

Suppose $F=(r,s)$ and $G=(t,v)$. There are two main cases, namely $r \neq t$, and $r=t$. Let $K=(w,z)$ so

$$w=-(t+r)+\lambda^2 \tag{1.7}$$

and

$$z=-v-\lambda(w-t), \tag{1.8}$$

where in the first case, we put

$$\lambda=\frac{s-v}{r-t}, \tag{1.9}$$

while in the second case, we put

$$\lambda=\frac{3r^2+A}{2s} \tag{1.10}$$

(I) Let us assume that we are in the case $r \neq t$ so λ is given by Equation (1.9).

(A) If $r(P)$ and $t(P)$ are finite and not equal, then $K(P)=F(P)+G(P)$ is simply the generic addition formula.

(B) If $r(P)$ and $t(P)$ are finite and equal, then we must have $s(P)=-v(P)$, or $s(P)=v(P)$.

(1) If $s(P)=-v(P)\neq 0$, then we have $F(P)=-G(P)$ therefore $F(P)+G(P)=\mathcal{O}$. On the other hand λ has a pole at P, therefore $K(P)=\mathcal{O}=F(P)+G(P)$.

(2) If $s(P) = v(P)$, since we have $r(P) = t(P) \neq \infty$, the formulas for $F(P)+G(P)$ and $K(P)$ only differ in the definition of λ. One has $v \neq -s$. Indeed if $v = -s$, $v(P) = s(P) = -s(P) = 0 = r(P)^3 + Ar(P) + B$. But one also has $r^3 + Ar + B = s^2 = v^2 = t^3 + At + B$, hence $(r-t)(r^2+rt+t^2+A) = 0$. Since $r \neq t$, and $r(P) = t(P)$, $3r^2(P)+A = 0$, which means that $r(P)$ is a double root of $X^3 + AX + B$; this is impossible since the curve equation is not singular. Hence we can write

$$\begin{aligned}\lambda &= \frac{v-s}{t-r} \cdot \frac{v+s}{v+s} \\ &= \frac{[t^3 - r^3] + A(t-r)}{(t-r)(v+s)} \\ &= \frac{t^2 + rt + r^2 + A}{s+v} \ .\end{aligned}$$

A. In the case $s(P) = v(P) \neq 0$ (and $r(P) = t(P)$), this becomes

$$\lambda(P) = \frac{3r(P)^2 + A}{2s(P)} \ ,$$

as required by Equation (1.10).

B. If $s(P) = v(P) = 0$, then we are at a point $F(P)$ of order two, *i.e.* $r(P)$ is a root of the curve equation. Hence $F(P) = G(P)$ has order two, and $F(P) + G(P) = \mathcal{O}$.

On the other hand we see from the expression of λ that its numerator is equal at P to $3r(P)^2 + A \neq 0$ (again because the curve equation is not singular) and therefore that λ has a pole at P. The components of $K(P)$ must also have poles since everything else besides λ in the addition formula is finite: $K(P) = \mathcal{O} = F(P) + G(P)$.

(C) Next we do the case where exactly one of $F(P)$ or $G(P)$, say $F(P)$, is $\mathcal{O}$. In this case, r and s have poles at P so we can write $r = r_1/u^d$ and $s = s_1/u^e$ where u is a uniformizing variable at P, d and e are positive integers, and r_1 and s_1 are rational functions that are finite and nonzero at P. Then since $s^2 = r^3 + Ar + B$, we must have $2e = 3d$ and $s_1^2(P) = r_1^3(P)$. The last statement follows from

$$\frac{s_1^2}{r_1^3} = 1 + \frac{Au^{2e}r_1 + Bu^{3e}}{r_1} \; .$$

Using Equations (7), (8), and (9), we compute w.

$$\begin{aligned}
w &= -(t+r) + \left(\frac{s-v}{r-t}\right)^2 \\
&= \frac{-(r^3 - t^2 r - t r^2 + t^3) + (s^2 - 2vs + v^2)}{(r-t)^2} \\
&= \frac{-r^3 + t^2 r + t r^2 - t^3 + (r^3 + Ar + B) - 2vs + v^2}{(r-t)^2} \\
&= \frac{t\,r^2 + (t^2 + A)r + (v^2 - t^3 + B - 2vs)}{r^2 - 2tr + t^2} \\
&= \frac{t\,r_1^2 u^{-2d} + (t^2 + A) r_1 u^{-d} + (v^2 - t^3 + B - 2v s_1 u^{-e})}{r_1^2 u^{-2d} - 3t\,r_1 u^{-d} + t^2} \\
&= \frac{t\,r_1^2 + u^d R_1 - 2v u^{2d-e} S_1}{r_1^2 + u^d R_2}\ ,
\end{aligned}$$

where R_1 and R_2 are rational functions that are finite at P. Since $2e = 3d, 2d - e > 0$, we see that $w(P) = t(P)$. This is what we want since $F(P) + G(P) = G(P)$ when $F(P) = \mathcal{O}$.

Now we could compute the y coordinate in a similar fashion, but we can avoid that by using the associative law on the elliptic curve of rational maps. Since $F+G = K$, we have $G - K = F$. Now the previous computation of the x coordinate shows that $K(P) \neq \mathcal{O}$, so we know that $(G-K)(P) = G(P) - K(P) = F(P) = \mathcal{O}$, which shows that $G(P) = K(P)$ as desired.

(D) The next case to consider is when F and G both have poles at P. Again we can use associativity of addition on the elliptic curve of rational maps to save a lot of work. Since $K = F + G$, $(F + G) - K = \mathcal{O}_M$, and by associativity $F + (G - K) = \mathcal{O}_M$. Suppose $K(P) = Q \neq \mathcal{O}$. Then since $G(P) = \mathcal{O}$, we can conclude that $(G - K)(P) = -Q$ by the previous case. Similarly we get that $(F+G-K)(P) = -Q$. This is impossible since $F + G - K = \mathcal{O}_M$. Therefore $K(P) = \mathcal{O}$ as desired.

(II) Now we suppose $r = t$, so that λ is given by equation (1.10) and $s = v$ or $s = -v$.

(a) If $s = -v$, then $F = -G$, so $K = \mathcal{O}_M$, the map whose constant value is $\mathcal{O}$. But we also have $F(P) = -G(P)$, so $F(P) + G(P) = \mathcal{O}$ for any point $P \in E$ which agrees with $K = \mathcal{O}_M$.

(b) If $s = v$, we have $F = G \neq \mathcal{O}_M$ and $K = 2F$.

If $r(P) \neq \infty$, then $K(P) = 2F(P)$ because they are identically defined. If $r(P) = \infty$, $F(P) = \mathcal{O}$ and the duplication formula yields

$$w = \frac{-4rs + 3r^2 + A}{2s} \ ,$$

which shows w has a pole at P if r and s do. As above, z must then also have a pole at P and $K(P) = \mathcal{O} = F(P) + G(P)$.

□

We now make an important definition.

Definition 1.7.2.

$$E[n] = \{P \in E : n \cdot P = \mathcal{O}\} \ .$$

Notice that $\mathcal{O} \in E[n]$ for all n, and, in fact, $E[n]$ is a subgroup of E. The points of $E[n]$ are called the *n-torsion* points of E.

Recall that g_n and h_n are defined by $n \cdot P = (g_n(P), h_n(P))$.

Theorem 1.7.3. *g_n and h_n are rational functions on E with poles precisely at the points of $E[n]$, and $E[n]$ has a finite number of points for all n.*

Proof. The proof is by induction on n. The functions $g_1(x, y) = x$ and $h_1(x, y) = y$ are clearly rational and nonzero, which gets the induction started. We now assume that g_n and h_n are rational for $n < q$ and that $E[n]$ is finite for $n < q$.

The idea of the inductive step is to write

$$q \cdot P = (q - 1) \cdot P + P. \tag{1.11}$$

By induction we can assume that g_{q-1} and h_{q-1}, the components of $(q - 1) \cdot P$, are rational functions of P. Our result will then follow from the previous theorem if we knew that $P \mapsto q \cdot P$ is not the rational map $\mathcal{O}_M$, *i.e.*, if we knew that $(q - 1) \cdot P \neq -P$ for some P. But $(q - 1) \cdot P = -P$ for all P means $E[q] = E$.

Suppose $E[q] = E$ and $k > 1$ divides q, and say $q/k = \ell$. Then $E[k]$ is a subgroup of $E[q]$, and if $P \in E[q], \ell \cdot P \in E[k]$. Since both k and ℓ are less than q, both $E[k]$ and $E[\ell]$ are finite. It is easy to see that if $E[\ell]$ is finite, then there are only finitely many points Q with $\ell \cdot Q = R$ for a fixed R. Therefore if q has a nontrivial divisor, $E[q]$ is finite, so we may as well assume that q is prime.

Observe that $E[2]$ is finite since there are only four points P with $2P = \mathcal{O}$, namely $\mathcal{O}$ and the three points of order two. Also $E[q] = E$ implies $E[2] \subset E[q]$, and $2P = \mathcal{O}$ and $qP = \mathcal{O}$ imply $P = \mathcal{O}$ if q is odd, therefore q is even. Since q is prime, this says $q = 2$, which contradicts the fact that $E[q]$ is infinite. Hence $E[q] \neq E$ and g_q is rational and not $\mathcal{O}_M$, so $E[q]$ is actually finite since it is the set of poles of a rational function not equal to $\mathcal{O}_M$. □

Corollary 1.7.4. *The rational function $g_n - x$ is not identically zero for any $n > 1$.*

Proof. If $g_n - x = 0$, then we would have $n \cdot P = \pm P$ or, equivalently, $(n \pm 1) \cdot P = \mathcal{O}$ for all $P \in E$. Hence either $E[n-1]$ or $E[n+1]$ would have to be infinite, contradicting the theorem. □

For the record, we write g_2 and h_2 here. Recalling that $s(x) = x^3 + Ax + B$, we have

$$g_2(P) = x(2 \cdot P) = \frac{x^4 - 2Ax^2 - 8Bx + A^2}{4s(x)}, \tag{1.12}$$

and

$$\begin{aligned} h_2(P) &= y(2 \cdot P) \\ &= y \; \frac{x^6 - 5Ax^4 + 20Bx^3 - 5A^2x^2 - 4ABx - 8B^2 - A^3}{8s(x)^2}. \end{aligned} \tag{1.13}$$

These formulas follow easily from the duplication formula.

In the next section, we will define a derivation of rational functions on E. Before we discovered this derivation, we used the following exercise to reduce to derivatives of functions of one variable. In the present treatment this fact is used very sparingly.

Exercise 1.7.5. Show that there are functions $\tilde{g}_n$ and $\tilde{h}_n$ of one variable such that $g_n(P) = \tilde{g}_n(x(P))$ and $h_n(P) = y(P)\tilde{h}_n(x(P))$ (Hint: Consider the mapping $F(P) = n \cdot P$. Show that $F(-P) = -F(P)$ and that the components of any map with this property must satisfy the exercise. Alternatively, do the exercise by induction.

We need to know some non-homogeneous information about g_n and h_n, *i.e.*, we would like to know their values at some point. Since the only point we can really get our hands on is $\mathcal{O}$, we examine the pole of g_n and h_n at $\mathcal{O}$. Let p be the characteristic of K.

Proposition 1.7.6. *If n is prime to p, then*

$$\frac{g_n}{x}(\mathcal{O}) = \frac{1}{n^2}$$

and

$$\frac{h_n}{y}(\mathcal{O}) = \frac{1}{n^3}\ .$$

Proof. Yet again, the proof is by induction on n. First assume that $n < p$. Our result clearly is true for $n = 2$ by Equations (12) and (13). Using the equation $n \cdot P = (n-1) \cdot P + P$, we get

$$\begin{aligned}\frac{g_n}{x} &= \frac{-g_{n-1}}{x} - 1 + \frac{1}{x}\left[\frac{h_{n-1} - y}{g_{n-1} - x}\right]^2 \\ &= \frac{-g_{n-1}}{x} - 1 + \frac{y^2}{x^3}\left[\frac{\frac{h_{n-1}}{y} - 1}{\frac{g_{n-1}}{x} - 1}\right]^2\end{aligned}$$

so

$$\begin{aligned}
\frac{g_n}{x}(\mathcal{O}) &= -(n-1)^{-2} - 1 + \left[\frac{(n-1)^{-3}-1}{(n-1)^{-2}-1}\right]^2 \\
&= (n-1)^{-2}\left[\frac{(n-1)^3-1}{(n-1)^2-1} - 1\right]\left[\frac{(n-1)^3-1}{(n-1)^2-1} + 1\right] - 1 \\
&= (n-1)^{-2}\left[\frac{(n-1)^3-(n-1)^2}{(n-1)^2-1}\right]\left[\frac{(n-1)^3-1}{(n-1)^2-1} + 1\right] - 1 \\
&= \frac{1}{n}\frac{(n-1)^2+2n}{n} - 1 \\
&= n^{-2}\ .
\end{aligned}$$

Similarly for h_n,

$$\begin{aligned}
\frac{h_n}{y} &= -\frac{h_{n-1}}{y} + \frac{1}{y}\frac{h_{n-1}-y}{g_{n-1}-x}(g_n - g_{n-1}) \\
&= -\frac{h_{n-1}}{y} + \frac{\frac{h_{n-1}}{y}-1}{\frac{g_{n-1}}{x}-1}\left(\frac{g_n}{x} - \frac{g_{n-1}}{x}\right)
\end{aligned}$$

$$\begin{aligned}
\frac{h_n}{x}(\mathcal{O}) &= -(n-1)^{-3} + \frac{(n-1)^{-3}-1}{(n-1)^{-2}-1}\left(n^{-2} - (n-1)^{-2}\right) \\
&= -(n-1)^{-3} + (n-1)^{-1}\frac{n^2-n+1}{n}(1-2n)n^{-2}(n-1)^{-2} \\
&= -(n-1)^{-3}n^{-3}\left(n^3 + (n^2-n+1)(1-2n)\right) \\
&= n^{-3}
\end{aligned}$$

It appears that we are stuck when the induction gets to p. Examining the previous computation, we see that

$$(g_{p-1}/x)(\mathcal{O}) = 1.$$

This is extremely unpleasant because $(g_{p-1}/x)(\mathcal{O}) - 1$ is in the denominator. It turns out, however, that there is an easy way around this problem that "bridges the gap" at the characteristic, but unfortunately gives no information about what happens at the characteristic.

The idea is to use the equation $n \cdot P = (n-2) \cdot P + 2 \cdot P$ to go directly from $p-1$ to $p+1$. The two relevant computations are similar to those above although they are longer. □

Corollary 1.7.7. *If n is prime to p, the leading coefficient of g_n is $1/n^2$ and the leading coefficient of h_n is $1/n^3$.*

Proof. The proposition shows that the order of g_n at $\mathcal{O}$ is two, so to compute the leading coefficient we must look at $(x/y)^2 g_n$. But

$$\left(\frac{x}{y}\right)^2 = \frac{x^2}{x^3 + Ax + B},$$

which has the same leading coefficient and order at $\mathcal{O}$ as $1/x$. Hence the first part of the corollary follows directly from the proposition. We leave the result about h_n as an exercise. □

Remark. This is a result that really depends on the characteristic. If n and p are *not* coprime, then the leading terms are quite different.

1.8 The Divisor of $g_m - g_n$

We want to compute the divisor of $g_m - g_n$ and relate it to the points of $E[k]$ for appropriate k. In order to compute the multiplicities of the zeros and poles of $g_m - g_n$, we must study the notion of derivative. It is possible to work with the functions $\tilde{g}_n$ and $\tilde{h}_n$ of one variable, so differentiation is just what we expect. It turns out that this is *not* the natural derivative on an elliptic curve, and the computations are unnecessarily complicated because of this. We want to define the derivative of an arbitrary rational function on E, but we must take care that the derivative of the polynomial $y^2 - x^3 - Ax - B$ is zero. If we formally take a derivative, we get

$$2yDy = (3x^2 + A)Dx,$$

which leads us to make the following definition:

Definition 1.8.1. Define a derivation D on the field of rational functions $K(E) = K(x, y)$ by setting

$$Dx = 2y$$

and

$$Dy = 3x^2 + A\ .$$

Extend D to arbitrary rational functions so that the usual rules of differentiation hold.

The following exercises should help to familiarize you with this notion:

Exercise 1.8.2.

(i) Show D is well-defined.

(ii) Suppose f is a nonzero polynomial and $f \neq g^p$ for any polynomial g. Show that the degree of Df (as a function of x and y) is one larger than the degree of f.

(iii) Let r be a rational function. Show that if r is finite at $P \in E$, then so is Dr. (Hint: You should handle the case $P = \mathcal{O}$ separately.)

Proof.

(i) D is first defined as a map from $K[x, y]$ to $K(E)$. Since $D(y^2 - s(x)) = 0$, it factors through $K(E)$ as a map from $K(E)$ to itself.

(ii) If $f(x, y) = a(x) + yb(x)$, $D(f) = 2y(a' + yb') + (3x^2 + A)b$, where for $h \in k[x]$, h' is the derivative of h in $k[x]$ and $\deg_x h$ is the degree of h as a polynomial in $k[x]$. As a consequence $\deg(Df) = \max(\deg(f) + 1, 4 + 2\deg_x b)$. But $\deg f \geq 3 + 2\deg_x b$, hence $\deg(Df) = \deg(f) + 1$.

(iii) First assume $P \neq \mathcal{O}$. write $r = f/g$ with f,g polynomials, g finite at P. $Df/g = Df.1/g + f.D1/g$ therefore it suffices to show that $D1/g$ is finite at P. But $D1/g = -1/g^2 Dg$, with both terms finite at P since Dg is a polynomial.

If $P = \mathcal{O}$, assume first that $\deg f < \deg g$. Then $\deg(Df.g - f.Dg) \leq 2\deg g$ and f/g is finite at P. On the other hand if $\deg f = \deg g$, write $f = a + by$, $g = c + dy$, a, b, c, d in $k[x]$. If $\deg a > \deg yb$, then $\deg a = \deg c = \deg f = \deg g$ and the leading terms of $Df.g$ and $f.Dg$ cancel out, so that

$\deg(Df.g - f.Dg) \leq 2\deg g$. If $\deg yb > \deg a$, then $\deg b = \deg d$; write $f/g = (ya+sb)/(yc+sd)$ to use the previous case again.

□

The basic fact we need to know about this derivative is the following:

Proposition 1.8.3. *Let r be a rational function, and a point P. If* $\operatorname{ord}_P(r) = d \neq 0$ *is prime to p, then* $\operatorname{ord}_P(Dr) = d - 1$. *If p divides d,* $\operatorname{ord}_P(Dr) \geq d$.

Proof. Suppose u is a uniformizing variable at P, and $r = u^d r_1$ where $r_1(P)$ is finite and nonzero. Then

$$Dr = du^{d-1}Du \cdot r_1 + u^d Dr_1 \ .$$

If we can show that Du is finite and nonzero at P, then we will have written $Dr = u^{d-1}r_2$ where r_2 is finite and nonzero at P, since the above exercise tells us that Dr_1 must be finite at P.

There are the usual three cases. If P is not $\mathcal{O}$ or of order two, then we can take $u(x,y) = x - x(P)$, so $Du = 2y$, which is finite and nonzero in this case. If P is a point of order two, we take $u = y$, and $Du = 3x^2 + A$. Since E is nonsingular, we know that the derivative of $f(x) = x^3 + Ax + B$ is nonzero at a zero of f. But $f'(x) = 3x^2 + A$ and if P has order two, $x(P) = \omega$, a zero of f. Hence Du is finite and nonzero at P in this case too.

If $P = \mathcal{O}$, we take $u = x/y$, and

$$Du = D(x/y) = \frac{2y^2 - 3x^3 - Ax}{y^2} = \frac{-y^2 + 2Ax + 3B}{y^2},$$

which is minus one at $\mathcal{O}$. □

The property of some particular choices of uniformizing variables at P that we used in the previous proof ($Du(P) \neq 0$) is in fact verified by all uniform variables:

Corollary 1.8.4. *A rational function u is a uniformizing variable at P if and only if it satisfies $u(P) = 0$ and $Du(P) \neq 0$.*

Proof. An uniformizing variable at P satisfies $\operatorname{ord}_P u = 1$ hence $\operatorname{ord}_P(Du) = 0$ by proposition 1.8.3, which means that $Du(P) \neq 0$.

Conversely, let u satisfy $u(P) = 0$ and $Du(P) \neq 0$ and let $d = \operatorname{ord}_P(u)$. Then $d \geq 1$; If $d \geq 2$ (prime to p or not), $\operatorname{ord}_P(Du) \leq 1$ and $Du(P) = 0$ which is impossible. Hence $d = 1$ and u is a uniformizing variable. □

We need the following proposition to compute the multiplicity of the zeros of $g_m - g_n$ and $h_m - h_n$.

Proposition 1.8.5. *We have $Dg_n = 2nh_n$ and $Dh_n = n(3g_n^2 + A)$.*

Proof. Again the proof is by induction on n. The case $n = 1$ is the definition of D. The case $n = 2$ follows upon differentiating Equation (1.12) and comparing it with n times Equation (1.13). For the inductive step, we have the following equations:

$$g_n = -g_{n-1} - x + \left(\frac{h_{n-1} - y}{g_{n-1} - x}\right)^2, \tag{1.14}$$

$$h_n = -y - \left(\frac{h_{n-1} - y}{g_{n-1} - x}\right)(g_n - x), \tag{1.15}$$

$$h_{n-1}^2 = g_{n-1}^3 + Ag_{n-1} + B, \tag{1.16}$$

and

$$Dh_{n-1} = (n-1)(3g_{n-1}^2 + A). \tag{1.17}$$

Equations (1.14) and (1.15) come from the generic addition formula applied to the equation $n \cdot P = (n-1) \cdot P + P$. Equation (1.16) simply expresses the fact that $(n-1) \cdot P = (g_{n-1}, h_{n-1})$ is on the curve E, and equation (1.17) is the inductive hypothesis.

Now we assume that $Dg_{n-1} = 2(n-1)h_{n-1}$ and differentiate (14) to get

$$\begin{aligned} Dg_n = -D_{g_{n-1}} - 2y + 2\left(\frac{h_{n-1} - y}{g_{n-1} - x}\right) \\ \left[\frac{(g_{n-1} - x)(Dh_{n-1} - Dy) - (h_{n-1} - y)(Dg_{n-1} - 2y)}{(g_{n-1} - x)^2}\right] . \end{aligned}$$

Then we use (17) to write the result in terms of powers of g_{n-1} and h_{n-1}. Next we use (16) to eliminate the powers of h_{n-1} greater than one obtaining

$$\begin{aligned} Dg_n = & \frac{1}{(g_{n-1} - x)^3} \\ & \{-2y + 2[(g_{n-1} - x)(-3x^3 + (n-1)(3g_{n-1^3} + A) - A) \\ & - (h_{n-1} - y)(2(n-1)h_{n-1} - 2y)]\} \\ & - 2(n-1)h_{n-1} \ . \end{aligned}$$

If we compare the result with nh_n using (15), we will see we get the same thing. □

This proposition is somewhat striking; it is not a result that was at all obvious from the equation of the curve or the addition formulas.

Before we get to the theorem about the divisor of $g_m - g_n$, we need a lemma about translations. This lemma will enable us to use information about the order of g_n at one point in $E[n]$ to get information about the order of g_n at all points of $E[n]$. A similar situation will arise in Section 13.

Lemma 1.8.6. *Let $P, Q \in E$, and suppose u is a uniformizing variable at P. Then the function $\mathcal{T}_Q(u)$ defined by*

$$[\mathcal{T}_Q(u)](R) = u(R+Q)$$

is a uniformizing variable at $P - Q$.

Proof. The point here is that $\mathcal{T}_Q$ is an automorphism of the field $K(E)$ so we can use $\mathcal{T}_{(-Q)}$ to go back. Suppose $\mathcal{T}_Q(u)$ is not a uniformizing variable at $P - Q$. Since it clearly has a zero at P_Q, it must have order $m > 0$. This means that $\mathcal{T}_Q$ takes a function of order d at P into a function of order md at $P - Q$. But this implies that $\mathcal{T}_{(-Q)}$ takes a function of order d at P_Q into a function of order d/m at P, which is absurd. □

It now follows that if a rational function r has divisor

$$\sum n(P) \langle P \rangle ,$$

then the function $\mathcal{T}_Q(r)$ has divisor

$$\sum n(P) \langle P - Q \rangle .$$

Recall that

$$E[n] = \{P \in E : n \cdot P = \mathcal{O}\} .$$

We write $\langle E[n]\rangle$ to denote the divisor whose nonzero entries are the points of $E[n]$, each with coefficient one.

At this point we begin to get into difficulty if the characteristic is 3. Hence from this point on we assume that the CHARACTERISTIC OF K IS NOT 3.

Theorem 1.8.7. *Suppose $m > n > 0$ and that $m, n, m-n$, and $m+n$ are all prime to p. Then*

$$\operatorname{div}(g_m - g_n) = \langle E[m+n]\rangle + \langle E[m-n]\rangle - 2\,\langle E[m]\rangle - 2\,\langle E[n]\rangle\,. \tag{1.18}$$

Proof. There are, as usual, various cases. First consider the points in both $E[n]$ and $E[m]$. By definition they are also in $E[m+n]$ and $E[m-n]$. Hence we must show that $g_m - g_n$ has order $1+1-2-2 = -2$ there. One of these points is $\mathcal{O}$. Proposition 1.7.6 tells us that both g_m and g_n have poles of multiplicity two there. Further Corollary 1.7.7 tells us that these poles cannot cancel out since g_m and g_n have different leading coefficients at $\mathcal{O}$, namely $1/m^2$ and $1/n^2$, and it is easy to see that our hypothesis implies that $m^2 \not\equiv n^2 (\operatorname{mod} p)$. Hence $g_m - g_n$ has a pole of order two at $\mathcal{O}$.

Notice that if $P \in E[n]$, then $\mathcal{T}_P(g_n) = g_n$ since $n\cdot(Q+P) = n\cdot Q$. Hence by Lemma 1.8.6, the order of g_n is the same at all points of $E[n]$. Thus we see that $g_m - g_n$ has order -2 at every point of $E[m]\cap E[n]$ as desired.

Second we consider the points that are in $E[m]$ but not in $E[n]$. These points are *not* in either $E[m+n]$ or $E[m-n]$. We must therefore show that $g_m - g_n$ has order -2 here also. Now g_n has order greater than zero here, and we have seen in the previous case that g_m has order -2 here so this case follows easily.

The third case is the difficult one. Here we consider the points that are neither in $E[m]$ nor $E[n]$. There are three subcases, P in $E[m-n]$ but not in $E[m+n]$, P in $E[m+n]$ but not in $E[m-n]$, and P in both $E[m+n]$ and $E[m-n]$. In each of these subcases , $mP = nP$ or $mP = -nP$ therefore $g_m - g_n$ has a zero at P; the problem is to determine the multiplicities. We will use the derivative for this. By Proposition 1.8.5, $D(g_m - g_n) = 2mh_m - 2nh_n$.

If P in $E[m-n]$ but not in $E[m+n]$, we have $m \cdot P = n \cdot P \neq -n \cdot P$, and since $P \notin E[m] \cup E[n]$, $m \cdot P \neq \mathcal{O}$ and $n \cdot P \neq \mathcal{O}$. In this case $h_m(P) = h_n(P)$, so $D(g_m - g_n)(P) = 2(m-n)h_m(P)$. Now $m-n$ is prime to p, and $h_m(P) \neq 0$ because $m \cdot P \neq -m \cdot P$ so $m \cdot P$ cannot be a point of order two. Hence $g_m - g_n$ has a simple zero here, which is what we want. The case P in $E[m+n]$ but not in $E[m-n]$ is symmetric and also works out as desired.

On the other hand, in the third subcase, we have both $m \cdot P = -n \cdot P$ and $m \cdot P = n \cdot P$ (and $m \cdot P \neq \mathcal{O}$ and $n \cdot P \neq \mathcal{O}$), and the situation is quite different. This case is equivalent to assuming that $2m \cdot P = 2n \cdot P = \mathcal{O}$. We see that $D(g_m - g_n)(P) = 0$ here, so the multiplicity is at least two. We must look at $DD(g_m - g_n)$. Proposition 1.8.5 tells us that

$$Dh_n = n(3g_n^2 + A) \ .$$

Since $2n \cdot P = \mathcal{O}$,we see that $n \cdot P$ is a point of order two. Therefore $g_n(P) = \omega$ and $h_n(P) = 0$. Hence

$$DD(g_m - g_n)(P) = (m^2 - n^2)(3\omega^2 + A) \ ,$$

which is nonzero since $m-n$ and $m+n$ are prime to p, and E is nonsingular. Thus $g_m - g_n$ has a double zero here as desired. □

Corollary 1.8.8. *If n is prime to p, then $E[n]$ has n^2 points.*

Proof. Let d_n be the number of points in $E[n]$. By taking degrees in Equation (1.18), we get $d_{m+n} + d_{m-n} - 2d_m - 2d_n = 0$. We know that $d_1 = 1$ and $d_2 = 4$. It is then easy to see that $d_n = n^2$ satisfies this recursion.

To see that this solution is unique, it suffices to take $\overline{d}_1 = 0$ and $\overline{d}_2 = 0$, and show that if $\overline{d}_n$ satisfies the recursion, $\overline{d}_n$ must also be 0 for all n prime to p. Let k be prime to p. If we take $m = k - 1$ and $n = 1$ in the recursion, we get $\overline{d}_k = 2\overline{d}_{k-1} - \overline{d}_{k-2}$, which says that $\overline{d}_k = 0$ for $k - 2$ and $k - 1$ prime to p. Now take $n = 2$ and $m = k - 2$. We get $\overline{d}_k = 2\overline{d}_{k-2} - \overline{d}_{k-4}$, which tells us that $\overline{d}_k = 0$ if $k - 4$ and $k - 2$ are prime to p. Finally take $n = 3$ and $m = k - 3$ to get $\overline{d}_k = 0$ if $k - 3$ and $k - 6$ are prime to p.

Suppose $k - 1$ is not prime to p. Then $k - 2$ must be prime to p, and if $k - 4$ were not prime to p, then we would have $p = 3$, which we have excluded.

Suppose $k - 2$ is not prime to p. Then $k - 3$ must be prime to p, and if $k - 6$ were not prime to p, then 4 would have to be a multiple of p, which is also excluded.

Hence as long as k is prime to p, we can find a case, which tells us that $\overline{d}_k = 0$. □

Exercise 1.8.9. Suppose n is prime to p. Show that $E[n]$ is isomorphic to $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$. (Hint: Use the fundamental theorem of Abelian groups).

Proof. The fundamental theorem for Abelian groups states in the finite case that every finite Abelian group is a direct product of terms of the form $\mathbb{Z}/q^\ell\mathbb{Z}$, and that this decomposition is unique up to the term order.

If q is a prime different from p, $E[q]$ has q^2 points, hence it is isomorphic either to $\mathbb{Z}/q^2\mathbb{Z}$ or to $(\mathbb{Z}/q\mathbb{Z})^2$. Since it does not have elements of order q^2, we are in the second case.

If q is a prime different from p and $\ell > 1$, assume the theorem is proven for $n = q^{\ell-1}$. Then consider the multiplication by q, $\varphi : E[q^\ell] \to E[q^{\ell-1}]$. φ has kernel $E[q]$ which has q^2 elements and is therefore surjective. Let $G = \varphi^{-1}(E[q^{\ell-1}])$. G is isomorphic to a product $\mathbb{Z}/q^{\ell_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/q^{\ell_k}\mathbb{Z}$. Such a product is transformed by φ into the product $\mathbb{Z}/q^{\ell_1-1}\mathbb{Z} \times \cdots \times \mathbb{Z}/q^{\ell_k-1}\mathbb{Z}$, hence $k = 2$ and $\ell_1 = \ell_2 = \ell$, $G = E[q^\ell]$.

If $n = rs$ with $r > 1$, $s > 1$ relatively prime and prime to p, assume the result is true for $E[r]$, $E[s]$. Then $E[n]$ has size n^2 and contains groups $E[r] = \mathbb{Z}/r^2\mathbb{Z}$, $E[s] = \mathbb{Z}/s^2\mathbb{Z}$. Since these groups have relatively prime sizes their intersection is $\{0\}$ and $E[n]$ contains a subgroup isomorphic to $\mathbb{Z}/r^2\mathbb{Z} \times \mathbb{Z}/s^2\mathbb{Z} \simeq \mathbb{Z}/n^2\mathbb{Z}$, but since this subgroup has n^2 elements it is actually equal to $E[n^2]$.

□

Remark. It is a fact (see the last remark of this part) that $E[p]$ is either $\{\mathcal{O}\}$ or $\mathbb{Z}/p\mathbb{Z}$. This shows that multiplication by the characteristic is quite different than multiplication by an integer prime to the characteristic.

1.9 The Division Polynomials

We would like to define a polynomial ψ_n that has divisor $\langle E[n]\rangle$. By Proposition 1.6.7, we can do this provided the sum of the points in $E[n]$ is $\mathcal{O}$ and the degree of $\langle E[n]\rangle$ is 0. If $P \in E[n]$ then $-P \in E[n]$, and if P is not a point of order two, P and $-P$ are distinct. Also if P is a point of order two, then all of the points of order two are in $E[n]$, and they all sum to zero. Hence the sum of the points in $E[n]$ is $\mathcal{O}$, but $\deg(\langle E[n]\rangle) = n^2$. Therefore if we let Δ be the divisor $\langle E[n]\rangle - n^2 \langle \mathcal{O}\rangle$, the sum of the points in Δ will still be zero, and $\deg(\Delta)$ will be 0 at least if n is prime to p. We will want to be able to compute the ψ_n's inductively so we will need to define them even if n is not prime to p. The way we are going to do this is to define them at first only in characteristic 0, prove what we want, then give a different definition for positive characteristic, and finally show that the results in characteristic zero imply the results in characteristic $p > 0$. Therefore until we say otherwise we assume that the CHARACTERISTIC OF K IS ZERO.

We now know that we can get a polynomial with the correct divisor, but it will not be unique. By Exercise 1.5.4, if we specify the leading coefficient we can select a unique one.

Definition 1.9.1. Let ψ_n be the unique polynomial whose divisor is $\Delta = \langle E[n]\rangle - n^2 \langle \mathcal{O}\rangle$ and whose leading coefficient is n.

Remark. Since the coefficient of $\mathcal{O}$ in the divisor Δ is $1 - n^2$, we see that the degree of ψ_n is $n^2 - 1$.

Exercise 1.9.2.

(i) Show that

$$\psi_n^2(P) = n^2 \prod_{P' \in E[n] - \mathcal{O}} \left[x(P) - x(P')\right] .$$

(Hint: Look at divisors and leading coefficients.)

(ii) Suppose n is odd. Show that ψ_n is a function of x alone and that its degree as a function of x is $(n^2 - 1)/2$.

(iii) Suppose n is even. Show that ψ_n is y times a function of x alone, and that the degree of this function of x is $(n^2-4)/2$.

Answers

(i) Since the divisor of $x - x(P')$ is $\langle P'\rangle + \langle -P'\rangle - 2\langle \mathcal{O}\rangle$, both sides of the equation have the same divisor. They also both have n^2 as leading coefficient, hence they are equal.

(ii) , (iii) If $\psi_n = f(x) + y\,g(x)$, $\psi_n^2(P) = (f^2 + y^2g^2) + 2y(f\,g)$. Because ψ_n^2 is a function of x only, f or g is 0. Let ω be a point of order 2. If $\psi_n = f(x)$, $\operatorname{ord}_\omega \psi_n$ is even; if $\psi = y\,g(x)$, $\operatorname{ord}_\omega \psi_n$ is odd. But when n is even, the points of order two are zeros of order 1 of ψ_n, hence $f = 0$, $\psi_n = y\,g(x)$ and the degree of g as a polynomial of $k(x)$ is $(n^2-4)/2$. if n is odd, points of order two are not zeros of ψ_n, hence $g = 0$ and $\psi_n = f(x)$, whose degree as a polynomial in $k(x)$ is $(n^2-1)/2$.

The goal of the rest of this section is to show that g_n and h_n can be computed in terms of the ψ_n's and that the ψ_n's satisfy a recursion that allows them to be computed.

Theorem 1.9.3. *Suppose $m > n > 0$. Then*

$$g_m - g_n = -\frac{\psi_{m+n}\psi_{m-n}}{\psi_m^2\psi_n^2} \tag{1.19}$$

Proof. By Theorem 1.8.7,

$$\operatorname{div}(g_m - g_n) = \langle E[m+n]\rangle + \langle E[m-n]\rangle - 2\langle E[m]\rangle - 2\langle E[n]\rangle\ .$$

By definition,

$$\operatorname{div}\left(\frac{\psi_{m+n}\psi_{m-n}}{\psi_m^2\psi_n^2}\right) = \langle E[m+n]\rangle + \langle E[m-n]\rangle - 2\langle E[m]\rangle - 2\langle E[n]\rangle\,.$$

The above two equations show that the two sides of Equation (1.19) have the same divisor.

Now g_n has leading coefficient $1/n^2$, while ψ_n has leading coefficient n. A brief computation shows that the two sides of Equation (1.19) have the same leading coefficient, which proves the theorem. □

Corollary 1.9.4. *For any $P \in E$*

$$g_n(P) = x(n \cdot P) = x(P) - \frac{\psi_{n+1}(P)\psi_{n-1}(P)}{\psi_n(P)^2} \tag{1.20}$$

Since $g_1 = x$, the proof is trivial. The next theorem gives us the basic properties of the division polynomials.

Theorem 1.9.5. *The polynomials ψ_n satisfy the following:*

(o) $\psi_0 = 0,$

(i) $\psi_1 = 1,$

(ii) $\psi_2(P) = 2y,$

(iii) $\psi_3(P) = 3x^4 + 6Ax^2 + 12Bx - A^2,$

(iv) $\psi_4(P) = 4y(x^6 + 5Ax^4 + 20Bx^3 - 5A^2x^2 - 4ABx - 8B^2 - A^3),$

(v) *for $m > n > 0$,*

$$\psi_n^2\psi_{m+1}\psi_{m-1} - \psi_m^2\psi_{n+1}\psi_{n-1} = \psi_{m+n}\psi_{m-n}. \tag{1.21}$$

Proof. (o) and (i) follow by definition. For (ii), note that $E[2]$ consists of $\mathcal{O}$ and the three points of order two. Also $\operatorname{div}(y) = \langle\omega_1\rangle + \langle\omega_2\rangle + \langle\omega_3\rangle - 3\langle\mathcal{O}\rangle = \langle E[2]\rangle - 4\langle\mathcal{O}\rangle$. Since the leading coefficient of $2y$ is manifestly 2, this proves (ii).

To do (iii), observe that by Corollary 1.9.4

$$x(2\cdot P) = x(P) - \frac{\psi_3(P)\psi_1(P)}{\psi_2(P)^2} \ .$$

We know $x(2\cdot P)$ (it is just $g_2(x)$, which is given by Equation (1.12), and we also know the other ψ's, so we can solve for ψ_3.

We can do a similar computation for ψ_4, but there is a better way which avoids computing g_3. Observe that a 4-torsion point P is either of order two or four. If P is of order two, then $y(P) = 0$. If P is of order four, then $2\ P$ is of order two, so $h_2(P) = 0$. A glance at Equation (1.13) yields ψ_4.

We have already done the hard part of (v) in Theorem 1.9.5. We write $g_m - g_n = (g_m - g_1) - (g_n - g_1)$, and using Equation (1.19) we get

$$-\frac{\psi_{m+n}\psi_{m-n}}{\psi_m^2\psi_n^2} = x - \frac{\psi_{m+1}\psi_{m-1}}{\psi_m^2} - x + \frac{\psi_{n+1}\psi_{n-1}}{\psi_n^2} \ ,$$

which proves the desired result. □

Corollary 1.9.6.

(i) For $k > 2$,

$$\psi_{2k} = \frac{\psi_k}{2y}\left(\psi_{k+2}\psi_{k-1}^2 - \psi_{k-2}\psi_{k+1}^2\right) \ .$$

(ii) For $k \geq 2$,

$$\psi_{2k+1} = \psi_{k+2}\psi_k^3 - \psi_{k+1}^3\psi_{k-1} \ .$$

Proof. For (i), take $m = k+1$ and $n = k-1$ in Equation (1.21). For (ii), take $m = k+1$ and $n = k$. □

We are left with one loose end to tie up, which we do in the next proposition.

Proposition 1.9.7. *If $P \in E$ and $n \geq 2$, then*

$$h_n(P) = y(n \cdot P) = \frac{\psi_{n+2}(P)\psi_{n-1}(P)^2 - \psi_{n-2}(P)\psi_{n+1}(P)^2}{4y\psi_n(P)^3} \quad (1.22)$$

Proof. The proof is again by induction. The beginning of the induction is easy. For the inductive step, we use

$$h_n = -y - \left(\frac{h_{n-1} - y}{g_{n-1} - x}\right)(g_n - x).$$

We know h_{n-1} in terms of the ψ's by induction, and we know g_n and g_{n-1} in terms of the ψ's by Corollary 1.9.4. After plugging these results in the above equation, simplifying and subtracting what we get from what we want, we will be left with showing

$$-\psi_{n-2}\psi_{n-1}^2\psi_{n+2} + \psi_{n-3}\psi_n^2\psi_{n+1} + \psi_2^2\psi_{n-1}^3\psi_{n+1} - \psi_2^2\psi_{n-2}\psi_n^3 = 0 \ .$$

This will follow by applying Equation (1.21) separately to the last two terms. □

We would now like to extend the results of this section to positive characteristic. Our argument is somewhat similar to the one that appears in [3] around page 39. The main idea is to note that the results we want are really polynomial identities in the ring $\mathbb{Z}[A, B, x, y]$, and hence still hold upon reduction mod p. It takes a bit of work to get this all together. From now on we let the CHARACTERISTIC OF K BE ARBITRARY > 3.

We use Theorem 1.9.5 now to *define* ψ_n.

Definition 1.9.8. The polynomials ψ_n are the unique polynomials that satisfy the following conditions:

(o) $\psi_0 = 0$,

(i) $\psi_1 = 1$,

(ii) $\psi_2(P) = 2y$,

(iii) $\psi_3(P) = 3x^4 + 6Ax^2 + 12Bx - A^2$,

(iv) $\psi_4(P) =$
$4y\,(x^6 + 5Ax^4 + 20Bx^3 - 5A^2x^2 - 4ABx - 8B^2 - A^3)$,

(v) for $k > 2$,

$$\psi_{2k} = \frac{\psi_k}{2y}\left(\psi_{k+2}\psi_{k-1}^2 - \psi_{k-2}\psi_{k+1}^2\right),$$

and for $k \geq 2$,

$$\psi_{2k+1} = \psi_{k+2}\psi_k^3 - \psi_{k+1}^3\psi_{k-1}. \tag{1.23}$$

Note that because of Theorem 1.9.3 and Corollary 1.9.6, this definition agrees with the one we have already in characteristic zero.

Now consider the field of rational functions in two indeterminates $\mathcal{A}$ and $\mathcal{B}$ over the field of rational numbers. Let E be the elliptic curve over this field with the defining polynomial

$$Y^2 = X^3 + AX + B. \tag{1.24}$$

In this case we can identify the rational functions on E with the quotient field of the ring $\mathbb{Z}[\mathcal{A}, \mathcal{B}, X, Y]$ subject to the relation (24). We can easily see that in this case the polynomials ψ_k are actually in the ring $\mathbb{Z}[\mathcal{A}, \mathcal{B}, X, Y]/(F(X, Y))$ where $F(X, Y) = Y^2 - X^3 - AX - B$. Now consider a polynomial identity such as (1.21), which is an identity in the ψ_k's involving only integer coefficients. We have proved this so far only for fields of characteristic zero. However we may now deduce that this identity will hold for the ψ_k's in an arbitrary elliptic curve as follows.

The ring $\mathbb{Z}[\mathcal{A}, \mathcal{B}, X, Y]$ has the property that there is a unique ring homomorphism from it to an arbitrary ring R where $\mathcal{A}, \mathcal{B}, X, Y$ are mapped to any elements of R. If we map these indeterminates to A, B, x and y in the function field of an arbitrary elliptic curve over any field, then the homomorphism induces a mapping from $\mathbb{Z}[\mathcal{A}, \mathcal{B}, X, Y]/(F(X, Y))$ into the function field of the curve. This mapping will obviously map the ψ's to the ψ's. It then follows that any identity that holds in characteristic zero and therefore in particular in $\mathbb{Z}[\mathcal{A}, \mathcal{B}, X, Y]/(F(X, Y))$ will also hold in any elliptic curve over any field.

Although polynomial identities can be transferred from $R = \mathbb{Z}[\mathcal{A}, \mathcal{B}, X, Y]/(F(X, Y))$ to the polynomial ring of any curve, it cannot always be done with identititities involving rational functions. Indeed, if $r/s = t/v$ with $r, s, t, v \in R$, which can be rewritten as $r\,v = s\,t$ in R, it may happen that s or v is equal to 0 as polynomials on some arbitrary curve, and the resulting relation is trivial. This is why some further work is needed to prove Corollary 1.9.4 and Proposition 1.9.7, *i.e.*, the expressions for g_n and h_n in terms of the ψ_n's, in characteristic p.

Theorem 1.9.9. *Let E be an elliptic curve on a field k of characteristic $p > 3$, $(g_n, h_n)(P) = n.P$ on E and ψ_n as defined in 1.9.8. Then*

(i) ψ_n is not identically zero for all $n > 0$.

(ii) For $n \geq 2$, g_n satisfies

$$g_n = x - \frac{\psi_{n+1}\psi_{n-1}}{\psi_n^2} \tag{1.25}$$

(iii) For $n \geq 2$, h_n satisfies

$$h_n = \frac{\psi_{n+2}\psi_{n-1}^2 - \psi_{n-2}\psi_{n+1}^2}{4y\psi_n^3} \tag{1.26}$$

Proof. It is easily seen that ψ_n is not identically zero for $n \leq 4$ and that equations (1.25) and (1.26) are satisfied for $n = 2$ (see equations (1.12), (1.13) and definition(1.9.8)). For $m \geq 3$ assume (1.25), (1.26) for $2 \leq n < m$, and that ψ_n is not identically zero all for $0 < n < m + 1$.

Now assume we are in characteristic zero for a moment. If we take the expression for g_m given by (1.25), and note that

$$g_m = -g_{m-1} - x + \left(\frac{h_{m-1} - y}{g_{m-1} - x}\right)^2 ,$$

then we can use Equations (1.25) and (1.26) for $n = m - 1$, (which we know are true in characteristic zero) to eliminate g_{m-1} and h_{m-1}. We thus get an identity in the ψ's alone.

Similarly starting with (1.23) and using

$$h_m = -y - \left(\frac{h_{m-1} - y}{g_{m-1} - x} \right) (g_m - x),$$

we get another identity in the ψ's alone.

What these identities are is not important; what is important is rather that they are polynomial identities in $\mathbb{Z}[\mathcal{A}, \mathcal{B}, X, Y]/(F(X, Y))$. Hence the preceding argument shows they hold in characteristic p. We would now like to convert these polynomial identities in characteristic p back into Equations (1.25) and (1.26). To do this we must divide by ψ_{m-2}, ψ_{m-1}, and ψ_m. Fortunately these are the ψ's that we have assumed are not identically zero in the inductive hypothesis. Hence (1.25) and (1.26) for $n = m$ hold in characteristic p.

Now we know that

$$g_m - x = \frac{\psi_{m+1}\psi_{m-1}}{\psi_m^2}. \tag{1.27}$$

By Corollary 1.7.4, we know that $g_m - x$ is not identically zero. Hence we get that $\psi_{m+1} \neq 0$, which allows us to continue the induction. □

Hence we have established Corollary 1.9.4 and Proposition 1.9.7 even in characteristic p.

About the only thing left to prove about the ψ_n's is that their divisor is $\langle E[n] \rangle - n^2 \langle \mathcal{O} \rangle$, which only holds for n prime to p. We have proved

$$g_n - x = -\frac{\psi_{n+1}\psi_{n-1}}{\psi_n^2}. \tag{1.28}$$

Since we know that $g_n - x$ only has poles on $E[n]$, we see that ψ_n has zeros on $E[n]$. Now we have to show that these zeros are simple and that there are no others. If n is prime to p, then we know that $\deg(\psi_n) = n^2 - 1$ because it is $n^2 - 1$ in characteristic zero, and if n is prime to p, the leading coefficient, n, does not reduce to zero. Since ψ_n has a pole at $\mathcal{O}$ and zeros on $E[n]$, it cannot have any other zeros because $E[n]$ has n^2 points (counting $\mathcal{O}$). Since the poles of $g_n - x$ on $E[n]$ have multiplicity two, Equation (1.25) shows that the zeros of ψ_n must be simple. This proves $\operatorname{div}(\psi_n) = \langle E[n]\rangle - n^2 \langle \mathcal{O}\rangle$ provided n is prime to p.

If n is not prime to p, we can still say something. Look at equation (1.23) for $k = n$:

$$\psi_{2n+1} = \psi_{n+2}\psi_n^3 - \psi_{n+1}^3\psi_{n-1}. \tag{1.29}$$

ψ_n must be prime to ψ_{n+1} because otherwise the above equation would imply that ψ_{2n+1} has a triple zero. Since $2n+1$ must be prime to p, this is impossible by the result of the previous paragraph. Similarly by looking at Equation (1.23) with $k = n-1$, we see that ψ_n must be prime to ψ_{n-1}. Hence Equation (1.25) implies that ψ_n has all of its zeros on $E[n]$ but we do not know they are simple. Thus we have proved

Proposition 1.9.10. *If n is prime to p,* $\operatorname{div}(\psi_n) = \langle E[n]\rangle - n^2 \langle \mathcal{O}\rangle$ *even in positive characteristics. Even if n is not necessarily prime to p, ψ_n has all of its zeros on $E[n]$.*

Remark. Our final remark concerns $E[p]$. In characteristic zero the leading coefficient of ψ_n is n. Hence in characteristic p the degree of ψ_p is less than $p^2 - 1$. Since the elements of $E\lceil p]$ are all zeros of ψ_p, this says that $E[p]$ cannot have p^2 elements. Since $E[p]$ is a group all of whose elements are p-torsion, $E[p]$ must either be the trivial group or $\mathbb{Z}/p\mathbb{Z}$. It turns out that both cases occur.

Chapter 2

Elliptic Curves over Finite Fields

2.1 Objective

We let p be a prime, n some integer, $q = p^n$, and $k = \mathrm{GF}(q)$, the field with q elements. Let K be the algebraic closure of k. In particular, K is infinite. In this part we are interested in elliptic curves defined over k, so the points of our curve E lie in $K \times K$ and satisfy the equation

$$Y^2 = X^3 + AX + B \tag{2.1}$$

for some fixed $A, B \in k$. Recall that $(a, b) \in E$ is k-rational if $a, b \in k$, and $E(k)$ is the set of k-rational points of E. We make the convention that $\mathcal{O}$ is k-rational so $E(k)$ has the structure of an elliptic curve itself. A natural question to ask is, how many points lie on $E(k)$? Another way of asking the same question is how many solutions does Equation (2.1) have in k. We will let E_q denote the number of k-rational points on E.

There is an heuristic argument that suggests that E_q is approximately $q+$ 1. Write $k^* = \{g, g^2, \ldots, g^{q-1} = 1\}$, so $(k^*)^2 = \{g^2, g^4, \ldots, g^{q-1} = 1\}$, and $|(k^*)^2| = (q-1)/2$. We might, therefore, expect that for about half of the elements $a \in k$, $a^3 + Aa + B$ will be a square. For each such a, there are two values, namely b and $-b$, with $b^2 = a^3 + Aa + B$. Therefore we might expect roughly $2\,q/2 = q$ finite solutions of (2.1), and, since $\mathcal{O}$ is k-rational, $q+1$ solutions altogether. We should remark here that there is no general formula known for E_q and that the best known algorithm ([5]) is somewhat complicated. The main theorem of this part will concern the difference between $q+1$ and E_q.

An important tool in the study of this problem is a rational mapping φ, called the Frobenius mapping. We need a trivial result before defining φ.

Exercise 2.1.1. If $(a, b) \in E$, then $(a^q, b^q) \in E$.

Proof. For every a in K, $(a+b)^p = a^p + b^p$ and $(a+b)^q = a^q + b^q$. Therefore for $s \in k[x]$, $s(a^q) = s(a)^q$ and if $b^2 = s(a)$, $(b^q)^2 = (b^2)^q = s(a)^q = s(a^q)$. □

Definition 2.1.2. The Frobenius mapping $\varphi : E \to E$ is defined by $\varphi(a, b) = (a^q, b^q)$.

Exercise 2.1.3. Show that $(a, b) \in E$ is k-rational if and only if $\varphi(a, b) = (a, b)$. (Hint: Look at the zeros of $x^q - x$.)

Proof. $x^q - x$ has at most q roots which include all of k hence k is exactly composed of the elements of K s.t. $x^q = q$. □

We state here the Main Theorem to be proved in the course of this part. It was conjectured by E. Artin and proved by H. Hasse. Generalizations of related results are known as the Weil Conjectures.

Theorem (Main Theorem, Hasse). *Let E be an elliptic curve defined over $k = \mathrm{GF}(q)$, and $t = q + 1 - E_q$. Then*

(i) $\varphi \circ \varphi - [t] \circ \varphi + [q] = \mathcal{O}_M$ *and*

(ii) $|t| \leq 2\sqrt{q}$,

where $[m]$ is the rational mapping $P \mapsto m \cdot P$.

Although the Main Theorem concerns elliptic curves over finite fields, many of the results leading to it are valid over any field. Only the results involving the Frobenius mapping require k to be finite. Therefore in the rest of this part, unless we say otherwise, k will be an arbitrary field of characteristic $\neq 2$ or 3, and K will be its algebraic closure.

2.2 The Ramification Index

This section is concerned with general rational mappings and contains results that could well have been done in Part I.

Lemma 2.2.1. *Suppose r is a rational function on E. If r is not constant, then r takes on all values (including ∞). Conversely, if r only takes a finite number of different values, it is constant.*

Proof. Lemma 1.4.10 and Theorem 1.4.7 show that if r is not constant, $r - a$ must have at least one zero and one pole for any $a \in K$, therefore a and ∞ are in the image of r. Conversely if the image of r is finite, it does not reach all of K since K is infinite, hence r is constant. □

Proposition 2.2.2. *A nonconstant rational mapping $F : E \to E$ is onto.*

Proof. Let $F = (r, s)$ where r and s are rational functions. If r were constant, then s would have only finitely many values, and hence by the lemma, s would be constant. Then F would be constant, which is a contradiction. Similarly we can see that s is not constant.

It follows that both r and s have poles, but these must occur simultaneously since $(r(P), s(P))$ is always on the curve. Thus F takes the value $\mathcal{O}$. To see that F takes the value $P \in E$, apply this result to the function $Q \mapsto F(Q) - P$. □

Now we define the ramification index of a nonconstant rational mapping $F : E \to E$ at a point. Let $P \in E$ and u be a uniformizing variable at $F(P)$. If $u \circ F$ were identically zero, then u would be zero on $F(E)$. Since the previous proposition shows $F(E) = E$, this would imply that u would be identically zero itself, which cannot be. Thus $u \circ F$ is zero at P, but not identically zero on E.

Definition 2.2.3. The ramification index of F at P is defined by

$$e_F(P) = \mathrm{ord}_P(u \circ F),$$

where u is a uniformizing variable at $F(P)$.

It is easily verified that $e_F(P)$ is independent of the choice of u. Note that $e_F(P) \geq 1$. The principal property of the ramification index is the following:

Proposition 2.2.4. *Suppose that r is a nonzero rational function on E and F is a nonconstant rational mapping. Let $P \in E$. Then*

$$\operatorname{ord}_P(r \circ F) = [\operatorname{ord}_{F(P)} r] \cdot [e_F(P)] \ .$$

Exercise 2.2.5. Prove this proposition. (Hint: Take uniformizing variables at P and $F(P)$ and write everything out.)

Proof. write $r = u^{\operatorname{ord}_{F(P)} r} s$ with u an uniformizing variable at $F(P)$, and s finite and nonzero at $F(P)$. Then

$$r \circ F = (u \circ F)^e (s \circ F)$$

with $s \circ F$ finite and nonzero at P, therefore $\operatorname{ord}_P(r \circ F) = \operatorname{ord}_{F(P)} r \cdot \operatorname{ord}_P(u \circ F) = \operatorname{ord}_{F(P)} r \cdot e_F(P)$. □

We now use the ramification index to investigate the effect of a rational mapping on divisors.

Definition 2.2.6. Suppose that F is a nonconstant rational mapping. We define $F^* : \operatorname{div}(E) \to \operatorname{div}(E)$ to be the homomorphism with

$$F^*(\langle Q \rangle) = \sum_{F(P)=Q} e_F(P) \langle P \rangle \ .$$

Proposition 2.2.7. *F^* is one-to-one.*

Exercise 2.2.8. Prove this proposition.

Proof. If a divisor d has nonzero coefficient at Q, $F^*(d)$ has nonzero coefficient at any P s.t. $F(P) = Q$, and is therefore nonzero. Therefore its kernel is trivial. □

The preceding definition is made so that the following is true:

Proposition 2.2.9. *Suppose that F is a nonconstant rational mapping and that r is a nonzero rational function. Then*

$$\operatorname{div}(r \circ F) = F^*(\operatorname{div}(r)) \ .$$

Proof. Like most of the proofs in this part, this one is a straightforward computation.

$$\begin{aligned}
\operatorname{div}(r \circ F) &= \sum_P \operatorname{ord}_P(r \circ F) \langle P \rangle \\
&= \sum_P [\operatorname{ord}_{F(P)} r] \cdot [e_F(P)] \langle P \rangle \\
&= \sum_Q \operatorname{ord}_Q(r) \cdot \sum_{F(P)=Q} e_F(P) \langle P \rangle \\
&= \sum_Q [\operatorname{ord}_Q r] \cdot F^*(\langle Q \rangle) \\
&= F^*(\operatorname{div}(r)) \ .
\end{aligned}$$

□

We also need

Lemma 2.2.10. *Suppose that F_1 and F_2 are nonconstant rational mappings. Then $F_1 \circ F_2$ is nonconstant and for $P \in E$,*

$$e_{F_1 \circ F_2}(P) = e_{F_1}(F_2(P)) \cdot e_{F_2}(P) \ .$$

Proof. Since rational mappings are onto, it follows that $F_1 \circ F_2$ is nonconstant. Let u be a uniformizer at $F_1(F_2(P))$. Then

$$\begin{aligned} e_{F_1 \circ F_2}(P) &= \operatorname{ord}_P(u \circ F_1 \circ F_2) \\ &= [\operatorname{ord}_{F_2(P)}(u \circ F_1)] e_{F_2}(P) \\ &= e_{F_1}(F_2(P)) \cdot e_{F_2}(P) \ . \end{aligned}$$

□

The next proposition justifies the use of the upper star.

Proposition 2.2.11. *Suppose that F_1 and F_2 are nonconstant rational mappings. Then*

$$(F_1 \circ F_2)^* = F_2^* \circ F_1^* \ .$$

Proof. The proof is again a routine computation.

$$\begin{aligned} F_2^* \circ F_1^*(\langle R \rangle) &= F_2^* \left(\sum_{F_1(Q)=R} e_{F_1}(Q) \, \langle Q \rangle \right) \\ &= \sum_{F_1(Q)=R} e_{F_1}(Q) \cdot \sum_{F_2(P)=Q} e_{F_2}(P) \, \langle P \rangle \\ &= \sum_{F_1 \circ F_2(P)=R} e_{F_1}(F_2(P)) \cdot e_{F_2}(P) \, \langle P \rangle \\ &= \sum_{F_1 \circ F_2(P)=R} e_{F_1 \circ F_2}(P) \, \langle P \rangle \\ &= (F_1 \circ F_2)^*(\langle R \rangle) \ . \end{aligned}$$

□

Proposition 2.2.12. *For $P \in E$, let $\mathcal{T}_P$ be the translation sending Q to $Q+P$. Then $\mathcal{T}_P$ has ramification index 1 at every point. More generally, any invertible map F from E to itself has ramification index 1 at every point.*

Proof. The general case is a consequence of lemma 2.2.10, applied to F and its inverse. The result for translations is also a rewording of lemma 1.8.6. □

2.3 Endomorphisms

We now study a special class of rational mappings that contains the Frobenius mapping.

Definition 2.3.1. A rational mapping from E to E that is also a group homomorphism is called an *endomorphism*. These mappings form a group, which we denote by $\operatorname{End}(E)$.

Remark. We could also study rational mappings between different elliptic curves and, in particular, those which are homomorphisms, but we do not need them for proving the Main Theorem. Many of the ideas presented here can be easily extended to homomorphisms between elliptic curves.

Example 2.3.2.

(i) The mapping $[m]$ defined by $[m](P) = m \cdot P$ is clearly an endomorphism.

(ii) The Frobenius mapping is an endomorphism.

The following is a striking and important result:

Theorem 2.3.3. *Suppose $\alpha : E \to E$ is a nonzero endomorphism. Then the ramification index $e_\alpha(P)$ is independent of P.*

Proof. For $P \in E$, let $\mathcal{T}_P$ be the translation sending Q to $Q + P$. Since α is an endomorphism, $\alpha \circ \mathcal{T}_P = \mathcal{T}_{\alpha(P)} \circ \alpha$. Applying Lemma 2.2.10 to both sides of this equation at the point $\mathcal{O}$ yields

$$e_\alpha(P) \cdot e_{\mathcal{T}_P}(\mathcal{O}) = e_{\mathcal{T}_{\alpha(P)}}(\alpha(P)) \cdot e_\alpha(\mathcal{O}) \ .$$

But since translations have ramification index one at every point (proposition 2.2.12), $e_\alpha(P) = e_\alpha(\mathcal{O})$.

□

If α is an endomorphism, we denote by e_α the constant value of $e_\alpha(P)$ for $P \in E$. Now we show how to find e_α in the cases of interest.

Lemma 2.3.4. *Let m be any integer, r any rational function, and D the derivation of Section 1.8. Then*

$$D(r \circ [m]) = (m \cdot Dr) \circ [m] \ .$$

Proof. This is obvious for $m = 0$. If $r = x$, then $r \circ [m] = g_m$, and the result follows from 1.8.5. Similarly if $r = y$, we are done. One checks immediately that the set of those rational functions for which the lemma holds is closed under field operations $(+, -, \times, \mathrm{div})$. This proves the lemma for $m \geq 0$. The case $m = -1$ is easily verified directly.

Now take $m \geq 0$. We get

$$\begin{aligned} D(f \circ [-m]) &= D(f \circ [m] \circ [-1]) \\ &= -D(f \circ [m]) \circ [-1] \\ &= (-m) Df \circ [-m] \ . \end{aligned}$$

□

Proposition 2.3.5. *Suppose that E is defined over $k = \mathrm{GF}(q)$ and that φ is the Frobenius mapping. Then $e_\varphi = q$.*

Proof. We know that $e_\varphi = e_\varphi(\mathcal{O}) = \operatorname{ord}_{\mathcal{O}} \varphi$. $u = x/y$ is a uniformizing variable at $\mathcal{O}$ and

$$u \circ \varphi = \left(\frac{x}{y}\right) \circ \varphi = \frac{x^q}{y^q} = u^q.$$

therefore $\operatorname{ord}_{\mathcal{O}} \varphi = q$. □

Definition 2.3.6. Let $y : E \to E$ be an endomorphism. If $e_\alpha = 1$, we say α is *separable*. If $e_\alpha > 1$, we say α is *inseparable*.

Remark. Let $F : E \to E$ be a rational function. We can define a map $F^* : K(E) \to K(E)$ by $F^*(r) = r \circ F$. Then $F^*(K(E))$ will be some subfield of $K(E)$. If F is a separable (respectively inseparable) endomorphism, then $K(E)$ is a separable (respectively inseparable) extension of $F^*(K(E))$. This result is not needed here and is proven in the second part of these notes (corollary 3.6.5).

The next result that we need is that the set of separable endomorphisms is closed under addition, but it takes a little work to get to it.

Lemma 2.3.7. *Suppose that r is a rational function of the single variable x and $r' = 0$ where r' is the usual derivative. Then $r(x) = \tilde{r}(x^p)$ for some rational function $\tilde{r}$.*

Proof. This is obvious for polynomials. Write $r = f/g$ with f and g relatively prime polynomials of the single variable x. Then $r' = 0$ implies $f'g = g'f$, but since f and g are relatively prime, we have $f|f'$ and $g|g'$. Hence f' and g' are zero, and f and g are functions of x^p. □

Proposition 2.3.8. *Suppose r is a rational function on E and $Dr = 0$. Then there is a rational function $\tilde{r}$ with $r(x, y) = \tilde{r}(x^p, y^p)$.*

Proof. First note that

$$y^p = y(y^2)^{\frac{p-1}{2}} = y \cdot s(x)^{\frac{p-1}{2}}$$

where $s(x) = x^3 + Ax + B$. Therefore r has a unique representation as

$$r(x, y) = u(x) + y^p v(x) \ ,$$

where u and v are rational functions of x alone. If $Dr = 0$, we have

$$[u'(x) + y^p v'(x)] \cdot 2y = 0 \ .$$

It follows that $u' = v' = 0$, and our result then follows from the previous lemma. □

Proposition 2.3.9. *Suppose α is an endomorphism. Then α is inseparable if and only if $D(r \circ \alpha) = 0$ for all rational functions r.*

Proof. If $D(r \circ \alpha) = 0$ for all rational functions r, then this is true in particular when $r = u$ is a uniformizing variable at $\alpha(P)$ for some $P \in E$. Hence $D(u \circ \alpha) = 0$, and by the previous proposition, $u \circ \alpha$ is a function of x^p and y^p. Because K is algebraically closed there exists p^{th}-roots in K and $t \in K(E)$ s.t. $u \circ \alpha = t^p$ therefore $u \circ \alpha$ must have order $e_\alpha = p \, \mathrm{ord}_P(t) > 1$ at P.

On the other hand, suppose there is a rational function r and a point $P \in E$ with

$$[D(r \circ \alpha)](P) \neq 0 \ .$$

Let $w = r - r(\alpha(P))$. Then $w \circ \alpha(P) = 0$ and

$$[D(w \circ \alpha)](P) = [D(r \circ \alpha)](P) \neq 0,$$

so $w \circ \alpha$ has a zero of multiplicity one at P. Thus

$$1 = \mathrm{ord}_P(w \circ \alpha) = [\mathrm{ord}_{\alpha(P)}(w)] \cdot e_\alpha$$

by Proposition 2.2.4, and we see that $e_\alpha = 1$. □

Corollary 2.3.10. *An endomorphism α is inseparable if and only if*

$$\alpha(x, y) = (u(x^p, y^p), v(x^p, y^p))$$

for rational functions u and v.

This follows immediately from the two previous propositions.

Proposition 2.3.11. *If α and β are inseparable endomorphisms, then so is $\alpha + \beta$.*

This follows immediately from Corollary 2.3.10.

Corollary 2.3.12. *If m is any integer prime to p, then $[m]$ is a separable endomorphism.*

Proof. Let $P \in E$ and let u be a uniformizer at $m \cdot P$. Then by Lemma 2.3.4,

$$\{D(u \circ [m])\}(P) = \{m \cdot Du\}(m \cdot P) \ .$$

which is nonzero; see corollary 1.8.4. The result then follows from proposition 2.3.9. □

Proposition 2.3.13. *Suppose that E is defined over $k = GF(q)$ and m and n are integers with m prime to p. If φ is the Frobenius endomorphism, then $[m] + [n] \circ \varphi$ is separable.*

Proof. Let $\alpha = [m] + [n] \circ \varphi$. If α is inseparable, then since $[m] = \alpha - [n] \circ \varphi$, $[m]$ would be the sum of two inseparable endomorphisms. The previous proposition then implies that $[m]$ would be inseparable, which contradicts Corollary 2.3.12. □

Recall that the kernel of an endomorphism α is

$$\{P \in E : \alpha(P) = \mathcal{O}\}.$$

Definition 2.3.14. Suppose that α is a nonzero endomorphism. Let $|\ker \alpha|$ denote the number of elements in the kernel of α. We define the *degree* of α by

$$\deg \alpha = |\ker \alpha| \cdot e_\alpha \ .$$

Remark. This is not the usual way the degree of a mapping is defined. For a rational mapping $F : E \to E$, we have seen that

$F^*(K(E))$ is a subfield of $K(E)$ (see the previous remark). It turns out that $K(E)$ is a finite-dimensional vector space over $F^*(K(E))$, and the dimension of $K(E)$ over $F^*(K(E))$ is the usual definition of the degree of F. This definition agrees with ours in the case of an endomorphism. These matters are discussed in more generality around page 76 in [6], where an endomorphism is called an isogeny. Since we do not need these more general notions, we omit them.

Exercise 2.3.15.

(i) If m is prime to $p, \deg([m]) = m^2$.

(ii) If E is defined over $\mathrm{GF}(q)$ (*i.e.* the curve equation has coefficients in $\mathrm{GF}(q)$), then the degree of the Frobenius endomorphism is q.

(iii) If α and β are nonconstant endomorphisms, then

$$\deg(\alpha \circ \beta) = (\deg \alpha) \cdot (\deg \beta) \ .$$

(iv) If α is a nonconstant endomorphism and $\Delta \in \mathrm{div}(E)$, then

$$\deg(\alpha^*(\Delta)) = (\deg \alpha) \cdot (\deg \Delta) \ .$$

Proof.

(i) $|\ker[m]| = m^2$ (corollary 1.8.8) and $e_m = 1$.

(ii) $e_\varphi = q$ and clearly $\ker \varphi = \{\mathcal{O}\}$: $|\ker \varphi| = 1$.

(iii) The ramification index is multiplicative. Since β is onto, for any P in $\ker \alpha$ there is $|\ker \beta|$ elements mapped by β to P, hence $|\ker \alpha \circ \beta| = |\ker \alpha|\, |\ker \beta|$.

(iv) Let $\Delta = \sum_Q n_Q \langle Q \rangle$. Then

$$\alpha^*(\Delta) = \sum_Q n_Q \sum_{P, F(P)=Q} e_\alpha(P) \langle P \rangle = e_\alpha \sum_Q n_Q \sum_{P, F(P)=Q} \langle P \rangle$$

and each inner term in the rightmost sum has degree $|\ker \alpha|$.

□

There is another result about α^* that is special to endomorphisms. Recall the map sum : $\operatorname{div}(E) \to E$, which takes the divisor $\sum n(P) \langle P \rangle$ into the point $\sum n(P) \cdot P$.

Proposition 2.3.16. *Let α be a nonzero endomorphism. For $P \in E$, pick P_0 with $\alpha(P_0) = P$. Then*

$$\operatorname{sum}[\alpha^*(\langle P \rangle) - \alpha^*(\langle \mathcal{O} \rangle)] = (\deg \alpha) \cdot P_0.$$

Proof. We have

$$\alpha^*(\langle P \rangle) = e_\alpha \sum_{\alpha(Q)=P} \langle Q \rangle = e_\alpha \sum_{\alpha(R)=\mathcal{O}} \langle P_0 + R \rangle \ .$$

Hence

$$\begin{aligned} \operatorname{sum}\left[\alpha^*(\langle P \rangle) - \alpha^*(\langle \mathcal{O} \rangle)\right] &= \operatorname{sum}\left[e_\alpha \sum_{\alpha(R)=O} (\langle P_0 + R \rangle - \langle R \rangle) \right] \\ &= e_\alpha \sum_{\alpha(R)=\mathcal{O}} P_0 \\ &= e_\alpha \, |\ker \alpha| \, P_0 \\ &= (\deg \alpha) P_0 \ . \end{aligned}$$

□

2.4 The Weil Pairing

Another important tool in our proof of the main theorem is the Weil pairing, which is a map from $E[m] \times E[m]$ to K. In order to define it, we will make frequent use of the result of proposition 1.6.7 that a divisor Δ is principal if and only if $\deg(\Delta) = 0$ and sum $(\Delta) = \mathcal{O}$.

Fix an integer m prime to p.

Lemma 2.4.1. *For $T \in E[m]$, the divisor $[m]^*(\langle T\rangle - \langle \mathcal{O}\rangle)$ is principal.*

Proof. Since $\deg(\langle T\rangle - \langle \mathcal{O}\rangle) = 0$,

$$\deg([m]^*(\langle T\rangle - \langle \mathcal{O}\rangle)) = \deg[m] \cdot \deg(\langle T\rangle - \langle \mathcal{O}\rangle) = 0$$

by (iv) of Exercise 2.3.15.

Now pick $T_0 \in E$ with $m \cdot T_0 = T$. Then

$$\mathrm{sum}([m]^*(\langle T\rangle - \langle \mathcal{O}\rangle)) = m^2 \cdot T_0$$

by Proposition 2.3.16, and $m^2 \cdot T_0 = m \cdot T = \mathcal{O}$. Therefore with $\Delta = [m]^*(\langle T\rangle - \langle \mathcal{O}\rangle)$, $\deg(\Delta) = 0$ and $\mathrm{sum}(\Delta) = \mathcal{O}$; by proposition 1.6.7, Δ is principal. □

One has

$$[m]^*(\langle T\rangle - \langle \mathcal{O}\rangle) = \sum_{R \in E[m]} \langle T_0 + R\rangle - \langle R\rangle .$$

Let g_T be a rational function with

$$\mathrm{div}(g_T) = [m]^*(\langle T\rangle - \langle \mathcal{O}\rangle) .$$

Although g_T is not unique, it *is* unique up to a constant multiple.

Let $\mathcal{T}_P$ be the translation by $P, i.e.\ \mathcal{T}_P(Q) = Q + P$.

Exercise 2.4.2. Show that $\mathcal{T}_P^*(\langle Q\rangle) = \langle Q - P\rangle$.

Proof. We know by proposition 2.2.12 that that translations have ramification index 1. The only preimage of Q by $\mathcal{T}_P$ is $Q-P$, hence the result. □

Lemma 2.4.3. *Suppose that $S, T \in E[m]$. Then*

$$\operatorname{div}(g_T \circ \mathcal{T}_S) = \operatorname{div}(g_T)\ .$$

Proof. Since $S \in E[m], [m] \circ \mathcal{T}_S = [m]$. Hence using Propositions 2.2.9 and 2.2.11, we get

$$\begin{aligned}
\operatorname{div}(g_T \circ \mathcal{T}_S) &= \mathcal{T}_S^* \circ [m]^*(\langle T\rangle - \langle \mathcal{O}\rangle) \\
&= ([m] \circ \mathcal{T}_S)^*(\langle T\rangle - \langle \mathcal{O}\rangle) \\
&= [m]^*(\langle T\rangle - \langle \mathcal{O}\rangle) \\
&= \operatorname{div}(g_T)\ .
\end{aligned}$$

□

The next proposition provides the basis for the definition of the Weil pairing.

Proposition 2.4.4. *Suppose $S, T \in E[m]$. Then the function $(g_T \circ \mathcal{T}_S)/g_T$ is constant, and its value is an m^{th}-root of unity in K and is independent of the choice of the function g_T.*

Proof. Since g_T is unique up to a constant multiple, it is clear that $(g_T \circ \mathcal{T}_S)/g_T$ does not depend on the choice of g_T.

By the lemma, there is an element $\zeta \in K$ such that $g_T \circ \mathcal{T}_S = \zeta g_T$. Composing this equation repeatedly with $\mathcal{T}_S$, we see that

$$g_T \circ \mathcal{T}_S^i = \zeta^i g_T\ .$$

Taking $i = m$, we see that $\zeta^m = 1$. □

Definition 2.4.5. Let $S, T \in E[m]$, and let μ_m be the group of m^{th}-roots of unity in K. Then the mapping from $E[m] \times E[m]$ to μ_m that sends (S, T) into $(g_T \circ \mathcal{T}_S)/g_T$ is called the *Weil pairing* and is denoted by w, *i.e.*,

$$w(S, T) = \frac{g_T \circ \mathcal{T}_S}{g_T} \ .$$

We summarize the properties of the Weil pairing in the following theorem:

Theorem 2.4.6. *Let S_1, S_2, S, T_1, T_2, $T \in E[m]$. Then the Weil pairing satisfies the following conditions:*

(i) $w(S_1 + S_2, T) = w(S_1, T) \cdot w(S_2, T)$.

(ii) $w(S, T_1 + T_2) = w(S, T_1) \cdot w(S, T_2)$.

(iii) $w(T, T) = 1$.

(iv) *If $w(S, T) = 1$ for all $S \in E[m]$, then $T = \mathcal{O}$.*

(v) *If α is any endomorphism, then*

$$w(\alpha(S), \alpha(T)) = w(S, T)^{\deg \alpha} \ .$$

Proof. (i) This is a straightforward computation.

$$\begin{aligned} w(S_1+S_2,T) &= \frac{g_T \circ \mathcal{T}_{S_1+S_2}}{g_T} \\ &= \frac{g_T \circ \mathcal{T}_{S_1} \circ \mathcal{T}_{S_2}}{g_T} \\ &= \left(\frac{g_T \circ \mathcal{T}_{S_2}}{g_T}\right) \circ \mathcal{T}_{S_1} \cdot \frac{g_T \circ \mathcal{T}_{S_1}}{g_T} \\ &= \frac{g_T \circ \mathcal{T}_{S_2}}{g_T} \cdot \frac{g_T \circ \mathcal{T}_{S_1}}{g_T} \\ &= w(S_2,T) \cdot w(S_1,T) \ . \end{aligned}$$

Where $\left(\frac{g_T \circ \mathcal{T}_{S_2}}{g_T}\right) \circ \mathcal{T}_{S_1} = \frac{g_T \circ \mathcal{T}_{S_2}}{g_T}$ because the latter is a constant function.

(ii) First note that

$$\operatorname{div}\left(\frac{g_{T_1+T_2}}{g_{T_1} \cdot g_{T_2}}\right) = [m]^*(\langle T_1+T_2\rangle - \langle T_1\rangle - \langle T_2\rangle + \langle \mathcal{O}\rangle) \ .$$

Now $\langle T_1+T_2\rangle - \langle T_1\rangle - \langle T_2\rangle + \langle \mathcal{O}\rangle$ is clearly principal. Let h be a function with this divisor, so

$$\operatorname{div}\left(\frac{g_{T_1+T_2}}{g_{T_1} \cdot g_{T_2}}\right) = [m]^*(\operatorname{div}(h)) = \operatorname{div}(h \circ [m]) \ ,$$

and we see that

$$\frac{g_{T_1+T_2}}{g_{T_1} \cdot g_{T_2}} = c \cdot h \circ [m]$$

for some $c \in K$. Hence if $S \in E[m]$,

$$\frac{g_{T_1+T_2}}{g_{T_1} \cdot g_{T_2}} \circ \mathcal{T}_S(P) = (c \cdot h \circ [m])(P+S) = (c \cdot h \circ [m])(P) \ ,$$

i.e., $\frac{g_{T_1+T_2}}{g_{T_1} \cdot g_{T_2}}$ is invariant under translation by elements of $E[m]$. Hence

$$\begin{aligned} w(S, T_1+T_2) &= \frac{g_{T_1+T_2} \circ \mathcal{T}_S}{g_{T_1+T_2}} \\ &= \frac{g_{T_1+T_2} \circ \mathcal{T}_S}{(g_{T_1} \circ \mathcal{T}_S)(g_{T_2} \circ \mathcal{T}_S)} \cdot \frac{(g_{T_1} \circ \mathcal{T}_S)(g_{T_2} \circ \mathcal{T}_S)}{g_{T_1+T_2}} \\ &= \frac{g_{T_1+T_2}}{g_{T_1} g_{T_2}} \cdot \frac{(g_{T_1} \circ \mathcal{T}_S)(g_{T_2} \circ \mathcal{T}_S)}{g_{T_1+T_2}} \\ &= \frac{g_{T_1} \circ \mathcal{T}_S}{g_{T_1}} \cdot \frac{g_{T_2} \circ \mathcal{T}_S}{g_{T_2}} \\ &= w(S, T_1) \cdot w(S, T_2) \end{aligned}$$

as desired.

(iii) This means that $g_T = g_T \circ \mathcal{T}_T$, *i.e.*, g_T is invariant by translation by T. Pick T_0 with $m \cdot T_0 = T$. Then using Propositions 2.2.9 and 2.2.11 and Exercise 2.4.2, and $[m] \circ \mathcal{T}_{i \cdot T_0} = \mathcal{T}_{i \cdot T} \circ [m]$, we get

$$\begin{aligned} \operatorname{div}(g_T \circ \mathcal{T}_{i \cdot T_0}) &= \mathcal{T}_{i \cdot T_0}^*(\operatorname{div} g_T) \\ &= \mathcal{T}_{i \cdot T_0}^* \circ [m]^*(\langle T \rangle - \langle \mathcal{O} \rangle) \\ &= ([m] \circ \mathcal{T}_{i \cdot T_0})^*(\langle T \rangle - \langle \mathcal{O} \rangle) \\ &= (\mathcal{T}_{i \cdot T} \circ [m])^*(\langle T \rangle - \langle \mathcal{O} \rangle) \\ &= [m]^* \circ \mathcal{T}_{i \cdot T}^*(\langle T \rangle - \langle \mathcal{O} \rangle) \\ &= [m]^*(\langle (1-i) \cdot T \rangle - \langle -i \cdot T \rangle) \ . \end{aligned}$$

It follows that the divisor of

$$G = g_T \cdot (g_T \circ \mathcal{T}_{T_0}) \cdot (g_T \circ \mathcal{T}_{2 \cdot T_0}) \cdots (g_T \circ \mathcal{T}_{(m-1) \cdot T_0})$$

is

$$\begin{aligned} &[m]^*[(\langle T \rangle - \langle \mathcal{O} \rangle) + (\langle \mathcal{O} \rangle - \langle -T \rangle) + (\langle -T \rangle - \langle -2 \cdot T \rangle) \\ &+ \cdots + (\langle (2-m) \cdot T \rangle - \langle (1-m) \cdot T \rangle)] \\ &= [m]^* \left[\langle T \rangle - \langle (1-m) \cdot T \rangle \right] \\ &= [m]^* \left[0 \right] = 0, \end{aligned}$$

since $T \in E[m]$. Therefore G is a constant, and when composed with $\mathcal{T}_{T_0}$, is the same constant. Hence

$$g_T \cdot (g_T \circ \mathcal{T}_{T_0}) \cdots (g_T \circ \mathcal{T}_{(m-1) \cdot T_0}) = \\ (g_T \circ \mathcal{T}_{T_0}) \cdot (g_T \circ \mathcal{T}_{2 \cdot T_0}) \cdots (g_T \circ \mathcal{T}_{m \cdot T_0}) \ ,$$

so after canceling, we get

$$g_T = g_T \circ \mathcal{T}_{m \cdot T_0} = g_T \mathcal{T}_T \ .$$

Thus $w(T,T) = g_T \circ a\mathcal{T}_T / g_T = 1$.

(iv) Suppose that $T \in E[m]$ and $w(S,T) = 1$ for all $S \in E[m]$. This means that $g_T \circ \mathcal{T}_S = g_T$ for all $S \in E[m]$, *i.e.*, g_T is invariant under translation by elements of $E[m]$. Because of lemma 2.4.7 below, $g_T = h \circ [m]$ for some rational function h. But then

$$[m]^*(\mathrm{div}(h)) = \mathrm{div}(g_T) = [m]^*(\langle T \rangle - \langle \mathcal{O} \rangle) \ .$$

By Proposition 2.2.7, we get $\langle T \rangle - \langle \mathcal{O} \rangle = \mathrm{div}(h)$ is a principal divisor. Lemma 1.4.10 then tells us that T must be $\mathcal{O}$.

(v) We need to show that

$$\left(\frac{g_T \circ \mathcal{T}_S}{g_T}\right)^{\deg \alpha} = \frac{g_{\alpha(T)} \circ \mathcal{T}_{\alpha(S)}}{g_{\alpha(T)}}.$$

But $\mathcal{T}_{\alpha(S)} \circ \alpha = \alpha \circ \mathcal{T}_S$, so if we compose the right side of the above equation with α leaving its constant value unchanged, we obtain

$$\frac{g_{\alpha(T)} \circ \alpha \ \circ \mathcal{T}_S}{g_{\alpha(T)} \circ \alpha}$$

Rewriting what we want to prove, we get

$$\frac{g_T^{\deg \alpha} \circ \mathcal{T}_S}{g_T^{\deg \alpha}} = \frac{g_{\alpha(T)} \circ \alpha \circ \mathcal{T}_S}{g_{\alpha(T)} \circ \alpha},$$

which is equivalent to showing

$$\left(\frac{g_{\alpha(T)} \circ \alpha}{g_T^{\deg \alpha}}\right) \circ \mathcal{T}_S = \frac{g_{\alpha(T)} \circ \alpha}{g_T^{\deg \alpha}},$$

i.e., we must show that

$$\frac{g_{\alpha(T)} \circ \alpha}{g_T^{\deg \alpha}}$$

is invariant under translation by elements of $E[m]$.

Since α is an endomorphism, α commutes with $[m]$, and thus α^* commutes with $[m]^*$, so

$$\begin{aligned} \operatorname{div}\left(\frac{g_{\alpha(T)} \circ \alpha}{g_T^{\deg \alpha}}\right) &= \alpha^* \circ [m]^*(\langle \alpha(T) \rangle - \langle \mathcal{O} \rangle) - \\ &\quad \deg(\alpha) \cdot [m]^*(\langle T \rangle - \langle \mathcal{O} \rangle) \\ &= [m]^*[\alpha^*(\langle \alpha(T) \rangle - \langle \mathcal{O} \rangle) - \\ &\quad \deg(\alpha) \cdot (\langle T \rangle - \langle \mathcal{O} \rangle)]. \end{aligned}$$

We show that the divisor Δ in square brackets is principal.

$\deg(\Delta) = \deg\alpha\ \deg(\langle\alpha(T)\rangle - \langle\mathcal{O}\rangle) - \deg\alpha\,(\langle T\rangle - \langle\mathcal{O}\rangle) = 0$;

by Proposition 2.3.16,

$$\mathrm{sum}(\alpha^*(\langle\alpha(T)\rangle - \langle\mathcal{O}\rangle)) = \deg(\alpha)\cdot T,$$

which cancels sum applied to the second term. Thus we have

$$\mathrm{div}\left(\frac{g_{\alpha(T)}\circ\alpha}{g_T^{\deg\alpha}}\right) = [m]^*(\mathrm{div}(h)) = \mathrm{div}(h\circ[m])$$

for some rational function h. It follows that

$$\frac{g_{\alpha(T)}\circ\alpha}{g_{T^{\deg\alpha}}}$$

is invariant under translation by elements of $E[m]$ as desired.

□

Lemma 2.4.7. *Suppose that r is a rational function on E that is invariant under translation by elements of $E[m]$. Then $r = t\circ[m]$ for some rational function t.*

The proof turns out to be surprisingly nontrivial.

Proof. Let H be the field of all rational functions on E that are invariant under translation by elements of $E[m]$. Let

$$J = \{g\circ[m] : g\in K(E)\}\ .$$

Then we certainly have $J\subset H\subset K(E)$, and we will show that $J = H$. We can consider each of these fields as a vector space over its subfield. Since H is the fixed field of a group of m^2 automorphisms ($E[m]$), Galois theory tells us the dimension of $K(E)$ over H is precisely m^2; see [2], chapter VI, theorem 1.8, or theorem 3.4.1 for a self-contained treatment.

We will show that $K(E)$ has dimension $\leq m^2$ when regarded as a vector space over J. This will show that $J = H$, and the lemma will follow.

Consider $J(x)$, the subfield of $K(E)$ generated by J and x. Recall the functions g_m and h_m from chapter 1. Since $g_m = x \circ [m]$ and $h_m = y \circ [m]$, we have $g_m, h_m \in J$. (In fact, J is generated by g_m and h_m.) By Exercise 1.7.5, $h_m = y\tilde{h}_m$ where $\tilde{h}_m$ is a function of x alone. Hence $y = h_m/\tilde{h}_m \in J(x)$. Hence $K(E) = J(x)$.

Now if we can show that x satisfies a polynomial in $J[X]$ of degree m^2, we will be done. Recall Equation (1.25),

$$g_m = x - \frac{\psi_{m-1}\psi_{m+1}}{\psi_{m^2}},$$

or

$$x\psi_m^2 - \psi_{m-1}\psi_{m+1} - \psi_m^2 g_m = 0 \ .$$

It follows from Exercise 1.9.2 (ii) and (iii) that ψ_m^2 and $\psi_{m-1}\psi_{m+1}$ can be written as polynomials in x alone of degree (in x) $m^2 - 1$ and m^2 respectively. Furthermore, since the leading coefficient of ψ_n is n, the coefficient of x^{m^2} in $x\psi_m^2 - \psi_{m-1}\psi_{m+1}$ is $m^2 - (m+1)(m-1) = 1$. Hence x satisfies the polynomial of degree m^2

$$X\psi_m^2(X) - \psi_{m-1}(X)\psi_{m+1}(X) - g_m\psi_m^2(X) = 0$$

in $J[X]$. □

We now give some corollaries of the theorem.

Corollary 2.4.8. *For $S, T \in E[m]$, $w(S,T) = w(T,S)^{-1}$.*

Proof. This follows from

$$w(S+T, S+T) = 1, w(S,S) = 1, w(T,T) = 1,$$

and (i) and (ii) of the theorem. □

Remark.

(i) The definition of the Weil pairing is dependent on m. let us denote by superscripts the choice of m. Then up to a constant, $g_T^{[m]} \circ [k] = g_T^{[m\times k]}$, so that for all $S, T \in E[m] \subset E[m \times k]$,

$$\begin{aligned} w^{[m]}(S,T) &= \frac{g_T^{[m]} \circ \mathcal{T}_S}{g_T^{[m]}} = \left(\frac{g_T^{[m]} \circ \mathcal{T}_S}{g_T^{[m]}} \right) \circ [k] \\ &= \frac{g_T^{[m\times k]} \circ \mathcal{T}_{S_0}}{g_T^{[m\times k]}} \\ &= w^{[m\times k]}(S_0, T) \end{aligned}$$

for $S_0 \in E[m\times k]$ s.t. $kS_0 = S$, because then $\mathcal{T}_S \circ [k] = [k] \circ \mathcal{T}_{S_0}$. As a result for all $T_0, S_0 \in E[m \times k]$,

$$w^{[m\times k]}(S_0, T_0)^k = w^{[m\times k]}(S_0, kT_0) = w^{[m]}(kS_0, kT_0)$$

so that the Weil pairing in $E[m]$ gives some partial information about the Weil pairing in $E[m \times k]$.

(ii) If m is not prime, then $\mathbb{Z}/m\mathbb{Z}$ is not a field, and $E[m]$ is not a vector space. $E[m]$ is, however, a free module of rank two over $\mathbb{Z}/m\mathbb{Z}$ (see exercise 1.8.9), so we can do linear algebra there.

Corollary 2.4.9. *Let T_1 and T_2 be a basis for $E[m]$ as a free module over $\mathbb{Z}/m\mathbb{Z}$. Then $w(T_1, T_2)$ is a primitive m^{th}-root of unity.*

Proof. Suppose $w(T_1, T_2)^n = 1$. Then $w(nT_1, T_2) = 1$. It then follows that $w(nT_1, c_1T_1 + c_2T_2) = 1$ for all $c_1, c_2 \in \mathbb{Z}$ from which we may conclude that $nT_1 = \mathcal{O}$ and m divides n. □

As our first application of the Weil pairing we present the following:

Theorem 2.4.10. *Suppose that α is a nonzero endomorphism. Then $\alpha(E[m]) \subset E[m]$. Furthermore, the determinant of α on $E[m]$ is equal* $\bmod\ m$ *to* $\deg(\alpha)$.

Proof. Let T_1 and T_2 be a basis for $E[m]$ over $\mathbb{Z}/m\mathbb{Z}$. Then for suitable integers $a_{i,j} \bmod m$ $(i, j = 1, 2)$, we have

$$\alpha(T_i) = \sum_{j=1}^{2} a_{i,j}T_j \ ,$$

and $\det(\alpha) = a_{1,1}a_{2,2} - a_{1,2}a_{2,1}$ as usual.

It is routine to check that this is independent of the choice of basis. We have by Theorem 2.4.6, (v)

$$w(\alpha(T_1), \alpha(T_2)) = w(T_1, T_2)^{\deg\alpha}$$

and

$$\begin{aligned} w(\alpha(T_1), \alpha(T_2)) &= w(a_{1,1}T_1 + a_{1,2}T_2, a_{2,1}T_1 + a_{1,2}T_2) \\ &= w(T_1, T_1)^{a_{1,1}\cdot a_{2,1}} \cdot w(T_1, T_2)^{a_{1,1}\cdot a_{2,2}} \\ &\quad \cdot w(T_2, T_1)^{a_{1,2}\cdot a_{2,1}} \cdot w(T_2, T_2)^{a_{1,2}\cdot a_{2,2}} \\ &= w(T_1, T_2)^{a_{1,1}a_{2,2} - a_{1,2}a_{2,1}} \\ &= w(T_1, T_2)^{\det\alpha} \ . \end{aligned}$$

Since $w(T_1, T_2)$ is a primitive m^{th}-root of unity (lemma 2.4.9), we are done. □

Remark. Note that $\det \alpha$ depends on m while $\deg \alpha$ is defined independently of m. The use of the Weil pairing allows us to pass from local information on $E[m]$ to global information on all of E.

Now we show that the degree of an endomorphism is essentially quadratic in the endomorphism. This will be important in proving the estimate of the number of k-rational points of E. First we need a lemma on 2×2 matrices.

Lemma 2.4.11. *Let A and B be 2×2 matrices with entries in some ring R. Then for $c_1, c_2 \in R$*

$$\begin{aligned} (i)\ \ &\det(c_1 A + c_2 B) \\ &\qquad = c_1^2 \det A + c_2^2 \det B + c_1 c_2 [\det(A+B) - \det A - \det B] \\ (ii)\ \ &\operatorname{tr} A \ \ = 1 + \det A - \det(I - A). \end{aligned}$$

The proof is an easy exercise.

Theorem 2.4.12. *If α and β are endomorphisms, then*

$$\deg(c_1\alpha + c_2\beta) = c_1^2 \deg \alpha + c_2^2 \deg \beta + c_1 c_2 [\deg(\alpha+\beta) - \deg \alpha - \deg \beta].$$

Proof. Let m be any integer prime to p. If we restrict α and β to $E[m]$, by Theorem 2.4.10 and the previous lemma we get mod m

$$\begin{aligned} \deg(c_1\alpha + c_2\beta) &\equiv \det(c_1\alpha + c_2\beta) \\ &\equiv c_1^2 \det \alpha \ + c_2^2 \det \beta + c_1 c_2 \\ &\qquad [\det(\alpha+\beta) - \det \alpha \ - \det \beta] \\ &\equiv c_1^2 \deg \alpha + c_2^2 \deg \beta + c_1 c_2 \\ &\qquad [\deg(\alpha+\beta) - \deg \alpha - \deg \beta]. \end{aligned}$$

The theorem now follows since this congruence holds for all m prime to p. □

The next theorem is the principal part of (i) of the Main Theorem.

Theorem 2.4.13. *If α is any endomorphism, then*

$$\beta = \alpha \circ \alpha - [1 + \deg \alpha - \deg(1-\alpha)] \circ \alpha - [\deg \alpha] = \mathcal{O} \ .$$

Proof. When we restrict to $E[m]$, we see that β becomes by lemma 2.4.11 (ii)

$$\alpha \circ \alpha - [1 + \det \alpha \ - \det(1-\alpha)] \circ \alpha - [\det \alpha] = \\ \alpha \circ \alpha - [\operatorname{tr} \alpha] \circ \alpha - \det \alpha.$$

But it is easy to see by direct computation that any 2×2 matrix A satisfies the equation

$$A^2 - (\operatorname{tr} A)A + \det A = 0 \ .$$

(This is also a special case of the Cayley-Hamilton Theorem.) Hence β restricted to $E[m]$ is zero. Since this holds for infinitely many m, β must be the zero endomorphism. □

Now we can prove the Main Theorem. Let E be defined over $k = \mathrm{GF}(q)$, and let φ be the Frobenius mapping. Let E_q be the number of k-rational points on E.

Theorem 2.4.14 (Hasse)**.** *Set $t = q + 1 - E_q$. Then*

(i) $\varphi \circ \varphi - [t] \circ \varphi + [q] = \mathcal{O}$ *and*

(ii) $|t| \leq 2\sqrt{q}$.

Proof. First note that $\ker(1-\varphi)$ is the set of $(a,b) \in E$ with $(a,b) = (a^q, b^q)$ together with the point $\mathcal{O}$. Exercise 2.1.3 tells us that $\ker(1-\varphi)$ is precisely the set of k-rational points of E, so $|\ker(1-\varphi)| = E_q$. Also Proposition 2.3.13 tells us that $1-\varphi$ is separable, so by our definition of degree, $\deg(1-\varphi) = E_q$. Since $\deg\varphi = q$ by (ii) of Exercise 2.3.15, (i) follows from Theorem 2.4.13.

To prove (ii), we note that

$$c_1^2 + c_2^2 q + c_1 c_2 (E_q - 1 - q) = \deg(c_1[1] - c_2\varphi) \geq 0$$

for all $c_1, c_2 \in \mathbb{Z}$. Hence

$$\left(\frac{c_1}{c_2}\right)^2 + q + \left(\frac{c_1}{c_2}\right)(E_q - 1 - q) = \left(\frac{1}{c_2}\right)^2 \deg\left(c_1[1] - c_2\varphi\right) \geq 0$$

for all rational numbers c_1/c_2. Thus we must have

$$v^2 + q + v(E_q - 1 - q) \geq 0$$

for all real numbers v. It follows that the discriminant of the quadratic function on the left must be ≤ 0. This discriminant is $t^2 - 4q$, which yields $|t| \leq 2\sqrt{q}$. □

Chapter 3

An Elementary Introduction to Elliptic Curves II

Introduction

Our goal in the previous chapters was to provide elementary proofs for the results used by Schoof in [5]. In doing this we left out those parts of the basic material on elliptic curves that were unnecessary for this purpose. In these notes we would like to fill in some of those gaps.

Our main topic concerns rational maps between elliptic curves. A secondary topic is Weil reciprocity which we use to further the study of the Weil pairing. Some of the results about rational maps between elliptic curves simply carry over from similar results proved previously about rational maps from an elliptic curve to itself. Many of the other results we present are usually proved using large amounts of commutative algebra. We have been able to find elementary proofs much in the style of [5]. A characteristic of these proofs is that although the theorems hold for more general varieties, our elementary proofs only work in the case of elliptic curves.

Unfortunately the situation is different in the case of Weil reciprocity. We have not been able to find a proof that avoids some more or less advanced notions, *i.e.*, Dedekind domains or valuation theory. We have chosen to present the required result without proof in the body of the notes and to relegate the more difficult material (and the proof) to a rather long Appendix. In this Appendix we develop the necessary theory starting from about the same level required in the rest of the notes so again you should not have to consult a lot of other books. One somewhat unsatisfactory result of this treatment is that the theorems of the Appendix hold in such generality that they include some of the earlier results.

We would like to thank David Robbins for insightful remarks, abundant useful criticism, and unstinting encouragement. We would also like to thank Noam Elkies for a number of incisive comments and suggestions.

In Section 3.1, we review some previous results about rational mappings and make the obvious extensions to rational mappings between elliptic curves. Section 2 provides a new proof of the key result that up to translation all rational mappings between elliptic curves are homomorphisms. Section 3 reviews some elementary field theory, while Section 4 is concerned with elementary Galois theory. Section 5 covers some elementary material on the norm map. Sections 3, 4, and 5 are required in the proof of the "Lower Star Theorem". They only contain standard results and can be omitted by a knowledgeable reader. Section 6 investigates the norm map on the various fields associated to an elliptic curve, and proves the “Lower Star Theorem”. Section 7 begins the discussion of Weil reciprocity. Section 8 contains standard results from valuation theory, and Section 9 completes the proof of generalized Weil reciprocity. Section 10 applies Weil reciprocity to yield a new expression for the Weil pairing.

We assume that our ground field has characteristic not equal to two or three.

3.1 Rational Maps Between Elliptic Curves

We let k be any field (with appropriate characteristic). We let K denote its algebraic closure. Let E and E' be elliptic curves over k with equations

$$Y^2 = X^3 + AX + B$$

and

$$Y^2 = X^3 + A'X + B'$$

respectively. We let $\mathcal{O}$ and $\mathcal{O}'$ denote the respective identity elements and x, y, x', and y' the coordinate functions.

Definition 3.1.1. A *rational map* F from E to E' is a pair (r, s) where r and s are rational functions on E such that

$$s^2 = r^3 + A'r + B' \ .$$

If we make the convention that $F(P) = \mathcal{O}'$ if and only if r and s are not finite at P, we see that F actually defines a map from E

to E' by $F(P) = (r(P), s(P))$ since r and s must have poles at the same points, and if they do not have poles, $(r(P), s(P))$ will be a point on E'.

Remark. There is an amusing way of looking at rational maps that will actually be useful. Given the field k, we form the elliptic curve E' using the equation

$$Y^2 = X^3 + A'X + B' \ . \tag{3.1}$$

Now suppose we consider the field of rational functions $K(E)$ on the curve E. Then we can use the same equation to form a new elliptic curve, which we might denote by $E'(K(E))$. Now $K(E)$ may not be algebraically closed, and by our convention, the points of $E'(K(E))$ have coordinates in the algebraic closure of $K(E)$. The finite points whose coordinates lie in $K(E)$ (*i.e.*, the $K(E)$-rational points) are precisely the rational maps from E to E'. We can think of the identity of this curve, call it $\mathcal{O}_M$, as the "map" with the constant value $\mathcal{O}'$.

All of the results and proofs on rational maps in chapter 2 carry over to the case of rational maps between elliptic curves with the obvious modifications. We state some here for ease of reference.

Proposition 3.1.2. *A nonconstant rational mapping $F : E \to E'$ is onto.*

Definition 3.1.3. The *ramification index of F at $P \in E$* is defined by

$$e_F(P) = \mathrm{ord}_P(u' \ \circ F),$$

where u' is a uniformizing variable at $F(P) \in E'$.

Proposition 3.1.4. *Suppose that r′ is a nonzero rational function on E' and $F : E \to E'$ is a nonconstant rational mapping. Let $P \in E$. Then*

$$\operatorname{ord}_P(r' \ \circ F) = [\operatorname{ord}_{F(P)} r'] \cdot [e_F(P)] \ .$$

Definition 3.1.5. Suppose that $F : E \to E'$ is a nonconstant rational mapping. We define $F^* : K(E') \to K(E)$ by $F^*(r') = r' \circ F$. We also define $F^* : \operatorname{div}(E') \to \operatorname{div}(E)$ to be the homomorphism with

$$F^*(\langle Q' \rangle) = \sum_{F(P)=Q'} e_F(P) \langle P \rangle \ .$$

Proposition 3.1.6. $F^* : \operatorname{div}(E') \to \operatorname{div}(E)$ *is one-to-one.*

Proposition 3.1.7. *Suppose that $F : E \to E'$ is a nonconstant rational mapping and that r′ is a nonzero rational function in $K(E')$. Then*

$$\operatorname{div}(F^*(r')) = \operatorname{div}(r' \circ F) = F^*(\operatorname{div}(r')) \ .$$

Proposition 3.1.8. *Suppose that $F_1 : E \to E'$ and $F_2 : E' \to E'$ are nonconstant rational mappings. Then*

$$(F_2 \circ F_1)^* = F_1^* \circ F_2^* \ .$$

3.2 Homomorphisms

In chapter 2 we studied *endomorphisms*, while here we study *homomorphisms.*

Definition 3.2.1. A rational mapping from E to E' that is also a group homomorphism is called a *homomorphism.* These mappings form a group, which we denote by $\operatorname{Hom}(E, E')$.

We are also interested in an apparently more general class of mappings.

Definition 3.2.2. A rational mapping $F : E \to E'$ with the property that $F(\mathcal{O}) = \mathcal{O}'$ is called an *isogeny*.

Clearly every homomorphism is an isogeny. The main result of this section is to show that every isogeny is a homomorphism.

Definition 3.2.3. A rational map $F : E \to E'$ is said to be *even* (respectively *odd*) if $F(-P) = F(P)$ (respectively $F(-P) = -F(P)$) for all $P \in E$.

It is obvious that an even homomorphism must be the map $\mathcal{O}_M$ that sends everything into $\mathcal{O}'$. If our main result is to hold, then *all* even maps must be constant since all maps are the composition of an isogeny and a translation. This is the first part of the proof of the main theorem of this section.

Theorem 3.2.4. *All even mappings are constant.*

Proof. Let $F : E \to E'$ be even. Write $r = x' \circ F$ and $s = y' \circ F$. Then $r, s \in k(E) = k(x, y)$. Now $F(-P) = F(P)$ implies $r(-P) = r(P)$ and $s(-P) = s(P)$, so by Exercise 1.7.5, $r, s \in k(x)$.

Hence it suffices to show that there are no nonconstant functions $r, s \in k(x)$ which satisfy

$$r^2 = s^3 + A's + B' \ .$$

This is really just a result about rational functions of one variable and has nothing to do with the curves E and E' *per se*. We can slightly restate it as follows: If d_1, d_2, d_3 are distinct in k, and $a, b \in k(X)$ satisfy

$$b^2 = (a - d_1)(a - d_2)(a - d_3), \tag{3.2}$$

then a and b are constant. (Here we use X to denote an indeterminate.)

We can write a and b in the form $a = q/c^2$ and $b = p/c^3$ for $p, q, c \in k[X]$ where we assume c has minimal degree. Then (3.2) becomes

$$(q - d_1c^2)(q - d_2c^2)(q - d_3c^2) = p^2. \tag{3.3}$$

So we want to show that (3.3) has no solutions (with d_1, d_2, d_3 distinct in k) such that q, c, and p are not all constants.

Now we show that the $(q - d_ic^2)$ are pairwise relatively prime for $i = 1, 2, 3$. Suppose that π were a common irreducible factor of $q - d_1c^2$ and $q - d_2c^2$. Then π divides q and c^2 and so divides $q - d_3c^2$ as well. Then (3) implies that $\pi^3|p^2$ so $\pi^4|p^2$. Thus π^2 divides one of the $q - d_ic^2$. Since $\pi|c^2$, we have $\pi^2|c^2$. Therefore we may conclude that $\pi^2|q$. It now follows that $\pi^6|p^2$ so that $\pi^3|p$. We could therefore replace p by p/π^3, q by q/π^2 and c by c/π to obtain another representation of a and b with a c of lower degree which is a contradiction.

Now we are reduced to showing that the system

$$\begin{aligned} q - d_1c^2 &= s_1^2 \\ q - d_2c^2 &= s_2^2 \\ q - d_3c^2 &= s_3^2 \end{aligned} \tag{3.4}$$

has no solutions where the s_i are pairwise relatively prime polynomials and q, c, s_1, s_2, and s_3 are not all constant. Subtracting and rewriting, we see that it suffices to show that the system

$$\begin{aligned} s_1^2 - s_3^2 &= t_1^2 c^2 \\ s_1^2 - s_2^2 &= t_2^2 c^2 \end{aligned} \tag{3.5}$$

has no solutions, s_1, s_2, s_3, which are pairwise relatively prime and s_1, s_2, s_3, and c are not all constants. In this system we are given that t_1^2 and t_2^2 (which are constants) are not equal or zero.

Assume we have such a solution and that $\max(\deg s_1, \deg s_3)$ is minimal among such solutions. Since s_1 and s_3 are relatively prime, (3.5) implies that we can write $s_1 - s_3 = 2f^2$ and $s_1 + s_3 = 2g^2$ for polynomials f and g which will be relatively prime and not both constant. (If f and g are both constant, then s_1 and s_3 will be constant. Hence c will be constant and so will s_2.)

We have

$$2\max(\deg f, \deg g) \leq \max(\deg s_1, s_3).$$

Since both f and g cannot be constant, $\max(\deg f, \deg g)$ is strictly less than $\max(\deg s_1, s_3)$. Using the second equation of (3.5), we see that

$$f^4 - \lambda f^2 g^2 + g^4 = s_{2^2} \tag{3.6}$$

and λ is not equal to ± 2 where $\lambda = 4(t_2/t_1)^2 - 2$.

If we factor (3.6), we get

$$(f^2 - \mu g^2)(f^2 - \mu' g^2) = s_{2^2} \tag{3.7}$$

where $\mu\mu' = 1$ and $\mu + \mu' = \lambda \neq \pm 2$. In particular we see that μ and μ' are not zero or equal. Hence there are polynomials h and k such that

$$\begin{aligned} f^2 - \mu g^2 &= h^2 \\ f^2 - \mu' g^2 &= k^2 \end{aligned} \tag{3.8}$$

and this system is of the same form as (3.5) with f corresponding to s_1 and h corresponding to s_3. Now (3.8) implies

$$\max(\deg f, \deg h) \leq \max(\deg f, \deg g)$$

which is strictly less than $\max(\deg s_1, \deg s_3)$. As this contradicts our assumption, we are done. □

Now it is not difficult to prove our main result.

Theorem 3.2.5. *All isogenies are homomorphisms.*

Proof. Let $\alpha_0 : E \to E'$ be an isogeny. Define $\alpha\pm : E \to E'$ by

$$\alpha_+(P) = \alpha_0(P) + \alpha_0(-P)$$
$$\alpha_-(P) = \alpha_0(P) - \alpha_0(-P)$$

for $P \in E$.

Then $2\alpha_0 = \alpha_+ + \alpha_-$. Since $\alpha_+(-P) = \alpha_+(P)$, we see by the previous theorem that α_+ is a constant map. Since

$$\alpha_0(\mathcal{O}) = \mathcal{O}', \quad \alpha_+(P) = \mathcal{O}' \quad \forall\, P \in E,$$

and we have

$$2\alpha_0 = \alpha_- \ .$$

Suppose we could show that α_- is a homomorphism. Then we would have

$$2\ (\alpha_0(P+Q) - \alpha_0(P) - \alpha_0(Q)) = \mathcal{O}'$$

which implies that

$$\alpha_0(P+Q) - \alpha_0(P) - \alpha_0(Q) \in E'[2]$$

for all $P, Q \in E$. Fixing Q, this is a nonsurjective map, and hence by Proposition 3.1.2 it is a constant function of P. Since α_0 is an isogeny, this constant must be $\mathcal{O}'$. Hence if we can show α_- is a homomorphism, then α_0 will also be one.

So now we can assume that we have an *odd* isogeny $\alpha : E \to E'$. Let $Q \in E$ and put

$$\beta_Q(P) = \alpha(P+Q) - \alpha(P-Q) \ .$$

It is now trivial to see that $\beta_Q(-P) = \beta_Q(P)$. Hence β_Q is constant, and we get

$$\beta_Q(P) = \beta_Q(\mathcal{O}) = \alpha(Q) - \alpha(-Q) = 2\alpha(Q) \tag{3.9}$$

for all $P \in E$.

Claim. $\alpha(n \cdot P) = n \cdot \alpha(P)$.

The proof of the claim is by induction. It is clear for $n = 0$ or 1. Consider

$$\begin{aligned}\beta_P(n \cdot P) &= \alpha((n+1) \cdot P) - \alpha((n-1) \cdot P) \\ &= \alpha((n+1) \cdot P) - (n-1) \cdot \alpha(P)\end{aligned}$$

where the last line follows by induction. On the other hand, by (3.9)

$$\beta_P(n \cdot P) = 2\alpha(P) \ .$$

Therefore

$$\alpha((n+1) \cdot P) = 2\alpha(P) + (n-1) \cdot \alpha(P) = (n+1) \cdot \alpha(P)$$

which proves the claim.

Put

$$\gamma_Q(P) = \alpha(P+Q) - \alpha(P) - \alpha(Q) \ .$$

Suppose $m \cdot P = Q$. Then

$$\begin{aligned}\gamma_Q(P) &= \gamma_{m\cdot P}(P) \\ &= \alpha(P + m\cdot P) - \alpha(P) - \alpha(m \cdot P) \\ &= (m+1)\cdot\alpha(P) - \alpha(P) - m\cdot\alpha(P) \\ &= \mathcal{O}' \ .\end{aligned}$$

Hence for fixed $Q, \gamma_Q(P) = \mathcal{O}'$ for all P such that $m \cdot P = Q$ for some m. Now it follows from corollary 1.8.8 that there are m^2 such P for each m prime to the characteristic of k. Since there are infinitely many such $P, \gamma_Q(P) = \mathcal{O}'$ for all $P, Q \in E$ which implies that α is a homomorphism. □

Henceforth we will use the words *isogeny* and *homomorphism* interchangeably. Notice that the theorem tells us that *any* rational mapping is the composition of an isogeny and a translation. This means that to assume that a rational mapping between elliptic curves is a homomorphism is not a very strong assumption.

3.3 Some Field Theory

Let $\alpha : E \to E'$ be a rational map, and let x' and y' be the coordinate functions on E'. Definition 1.3 defines a map $\alpha^* : \mathrm{div}(E') \to \mathrm{div}(E)$. It is easy to define a map in the other direction.

Definition 3.3.1. Define $\alpha_* : \mathrm{div}(E) \to \mathrm{div}(E')$ by setting

$$\alpha_*(\langle P \rangle) = \langle \alpha P \rangle$$

for the generators of $\mathrm{div}(E)$ and extending linearly.

One of the main objectives of this part of these notes is to show that if D is a principal divisor then so is $\alpha_*(D)$. In fact, we will define a map, also denoted by α_*, from $K(E)$ to $K(E')$ such that

$$\alpha_*(\mathrm{div}(r)) = \mathrm{div}(\alpha_*(r)) \ .$$

Note that if $\alpha(\mathcal{O}) = P'$, then $\beta = \mathcal{T}_{-P'} \circ \alpha$ takes $\mathcal{O}$ into $\mathcal{O}'$, and β is an isogeny. Furthermore if $D \in \mathrm{div}(E)$ and $\beta_*(D) = \mathrm{div}(r')$, then $\alpha_*(D) = \mathrm{div}(r' \circ \mathcal{T}_{P'})$. Hence it suffices to investigate the case of α a homomorphism.

It turns out that to study this situation (and, in fact, many others) it is advantageous to adopt the somewhat more abstract point of view alluded to in the remark on page 18 (see next paragraph). We will do that in this section. Most of the material is perfectly standard results from the elementary theory of fields, and so we will not give precise proofs for all of it. Good references are [7] and [19], Chapter II.

Let $K' = \alpha^*(K(E')) \subset K(E)$. K' consists of all functions in $K(E)$ of the form $r' \circ \alpha$ for some $r' \in E'$. Let $\overline{x} = x' \circ \alpha$ and $\overline{y} = y' \circ \alpha$ so $\overline{x}, \overline{y} \in K' \subset K(E)$. The idea here is to regard $K(E)$ as an extension of K'.

We show that $K(E)$ is a finite algebraic extension of K'. First observe that K' is isomorphic to $K(E')$ and hence is generated by $\overline{x}$ and $\overline{y}$ over K. Since y (respectively $\overline{y}$) satisfies an algebraic (in fact quadratic) equation over $K(x)$ (respectively $K(\overline{x})$), both $K(E)$

and K' have the same transcendence degree (namely 1) over K, *i.e.*, both $K(E)$ and K' have elements, x and $\overline{x}$ respectively, that do not satisfy any polynomial in $K[X]$ and such that $K(E)$ and K' are algebraic over $K(x)$ and $K(\overline{x})$ respectively.

Now if x were transcendental over K', then x and $\overline{x}$ would both be elements of $K(E)$ transcendental over K which are algebraically independent. This would mean that the transcendence degree of $K(E)$ over K would have to be at least two. Hence x is algebraic over K' and $K(E)$ is a finite algebraic extension of K'.

Recall that we said that α was *separable* if $e_\alpha = 1$. We want to see how this condition is reflected in the extension $K(E)$ over K'. We first need an easy and well-known lemma.

Lemma 3.3.2. *Let M be an extension of a field L and $r \in M$ be algebraic over L. Let f be a polynomial in $L[X]$ of minimal degree with $f(r) = 0$. Then f is irreducible, and if $g \in L[X]$ satisfies $g(r) = 0$, then f divides g. Hence for $r \in M$ algebraic over L there is a unique monic (i.e., leading coefficient one) irreducible polynomial $f \in L[X]$ with $f(r) = 0$.*

Exercise 3.3.3. Prove this lemma.

Proof. The set of polynomials cancelling on r is an ideal I of $L[X]$, generated by a minimal-degree polynomial f: $I = (f)$. If f were not irreducible, there would exist g and h such that $f = gh$ and $g(r) \neq 0$, $h(r) \neq 0$. But then since L is a field $f(r) \neq 0$, a contradiction. □

Definition 3.3.4. The unique monic irreducible polynomial $f \in L[X]$ with $f(r) = 0$ is called the *minimal* polynomial of r. We denote it by m_r.

Definition 3.3.5. We say an irreducible polynomial is *separable* if it has nonzero derivative. We say a reducible polynomial is *separable* if each of its irreducible factors is separable. A polynomial that is not separable is said to be *inseparable.*

If M is an extension of L, we say $r \in M$ is *separable over* L if m_r is separable. Otherwise we say r is *inseparable.* We say M is a *separable* extension of L if m_r is separable for each $r \in M$. Otherwise we say M is *inseparable* over L.

It is easy to see that an irreducible polynomial f is inseparable if and only if $f(X) = g(X^p)$ for some polynomial g where p is the characteristic of L. Suppose $f \in L[X]$ is inseparable so $f(X) = f_1(X^p)$ and the degree of f is divisible by p. If now f_1 is inseparable, then $f_1(X) = f_2(X^p)$ and the degree of f is divisible by p^2. Since the degree of f is finite, there is $s \geq 0$ such that $f(X) \in L[X^{p^s}]$, but $f \notin L[X^{p^{s+1}}]$. So we can write

$$f(X) = f_0\left(X^{p^s}\right),$$

and if $\deg f = n$ and $\deg f_0 = n_0$, then

$$n = n_0\, p^s \ .$$

Definition 3.3.6. The integer n_0 is called the *degree of separability* of f, while p^s is called the *degree of inseparability* of f.

It is clear that if $p = 0$ then everything must be separable.

Exercise 3.3.7. Show that if L is perfect (*i.e.*, of characteristic 0 or of characteristic $p > 0$ with every element of L a p-th power), then every polynomial in $L[X]$ is separable.

Proof. Let f be an irreducible polynomial in $L[X]$. In characteristic 0, f' is nonzero and theorefore coprime with f. In characteristic $p > 0$, if $f' = 0$, $f = f_0(X^p) = g_0(X)^p$ because every coefficient of f_0 has a p-th root: this is impossible because then f is not irreducible. Therefore $f' \neq 0$ and f is also separable in this case. □

One can prove the converse of this exercise (in the case of positive characteristic) if one knows the following result (which we need later anyway):

Lemma 3.3.8. *Let L be a field of characteristic p. Suppose $r \in L$ and there is no $s \in L$ with $s^p = r$. Then $X^{p^e} - r$ is irreducible in $L[X]$ for all $e \geq 0$.*

Proof. Let ρ be a root of φ in some extension L' of L. Then $r = \rho^{p^e}$ and

$$X^{p^e} - r = X^{p^e} - \rho^{p^e} = (X - \rho)^{p^e} \ .$$

Let $\varphi(X)$ be a nonconstant irreducible monic factor of $X^{p^e} - r$ in $L[X]$: because of the above, there exists k s.t. $\varphi = (X - \rho)^k$, and $X^{p^e} - r$ is a power of φ, say φ^m. Because $\deg(X^{p^e} - r)$ is a power of p, m must be a power of p. Hence $X^{p^e} - r = \varphi(X)^{p^t}$ for some $t \geq 0$. Let $c \in L$ be the constant term of φ. We have $r = (\pm c)^{p^t}$. Since r does not have a p-th root in L, we must have $t = 0$, so $X^{p^e} - r = \varphi(X)$ is irreducible in $L[X]$. □

Note that in characteristic $p > 0$, there are only two cases: either an element r does not have a p^{th}-root, or it has exactly one such root ρ. The corresponding polynomial $X^p - r$ then has ρ as a multiple root.

Exercise 3.3.9.

(i) Suppose L is a field of characteristic $p > 0$ and that every polynomial (of positive degree) in $L[X]$ is separable. Show that L is perfect.

(ii) Find a proof of the above lemma that does not use the fact that $X^{p^e} - r$ has a root in some extension. (Hint: Write $X^{p^e} - r = \varphi(X)\,\psi(X)$ for appropriate φ and ψ and differentiate.)

Proof.

(i) If L is not perfect then by lemma 3.3.8 a polynomial of the form $X^{p^e} - r$ is irreducible. It is not separable, which contradicts the hypothesis.

(ii) Assume $e \geq 1$ (otherwise there is nothing to prove) and $X^{p^e} - r = \varphi(X)\,\psi(X)$ with φ and ψ coprime, and $\deg \varphi \geq 1$. By differentiating, one gets

$$\varphi'\psi + \varphi\psi' = 0,$$

hence $\varphi'\psi = -\varphi\psi'$ and φ divides φ', ψ divides ψ': $\varphi' = \psi' = 0$. There are coprime φ_1 and ψ_1 s.t. $\varphi(x) = \varphi_1(X^p)$ and $\psi(X) = \psi_1(X^p)$;

$$X^{p^{e-1}} - r = \varphi_1(X)\,\psi_1(X)\ .$$

Iterating the process, we coprime find polynomials φ_e and ψ_e such that

$$X - r = \varphi_e(X)\,\psi_e(X)\ .$$

Then $\varphi_e = X - r$ and $\psi_e = 1$, which shows that $X^{p^e} - r$ is irreducible.

□

Let M be an extension of L. We know that $r \in M$ is inseparable if $m_r'(X) = 0$. It turns out, however, that it suffices merely to have m_r' vanish at r.

Proposition 3.3.10. *Suppose $r \in M$. Then r is inseparable over L if $m_r'(r) = 0$. Furthermore if r is inseparable over L and $g \in L[X]$ satisfies $g(r) = 0$, then $g'(r) = 0$.*

Proof. $m_r'(r) = 0$ implies $m_r' = 0$ since m_r' has lower degree than m_r. Hence r is inseparable. The converse is obvious.

If $g(r) = 0$, then $g(X) = h(X) \cdot m_r(X)$. Hence $g'(r) = h(r) \cdot m_r'(r)$ since $m_r(r) = 0$. If r is inseparable, we get $g'(r) = 0$ as desired. □

Now suppose $g \in L[X]$ and $g(r) = 0$. Then $X - r$ divides g considered as a polynomial in $M[X]$, *i.e.*, we can write

$$g(X) = (X - r)^s g_1(X)$$

for some integer $s \geq 1$ and some $g_1(X) \in M[X]$ with $g_1(r) \neq 0$. Since r is a member of $L(r) \subset M$, we can apply the same argument to the field $L(r)$ instead of M. We get

$$g(X) = (X - r)^\sigma g_2(X)$$

for some integer $\sigma \geq 1$ and some $g_2(X) \in L(r)[X]$ with $g_2(r) \neq 0$. The above two equations imply that $\sigma = s$ and $g_1 = g_2$. Thus s depends only on $g(X)$ and the element r and not on the extension M.

Definition 3.3.11. We call s the *multiplicity* of the root r of g.

We have the following easy corollary which is just a restatement of the proposition.

Corollary 3.3.12. *Suppose $r \in M$. Then r is inseparable over L if and only if r is a multiple root of m_r. Furthermore if r is inseparable over L and $g(X) \in L[X]$ satisfies $g(r) = 0$, then r is a multiple root of g.*

Actually we can describe the multiplicity of the roots of an irreducible polynomial that splits completely in some extension. Suppose $f \in L[X]$ and M is an extension of L so that f factors completely in $M[X]$, *i.e.*,

$$f(X) = c_0(X - r_1)(X - r_2) \cdots (X - r_n)$$

with $r_i \in M$.

Proposition 3.3.13. *If $f(X) \in L[X]$ is irreducible and has degree of inseparability p^s, then each linear factor in the above equation appears exactly p^s times.*

Proof. We have $f(X) = g\left(X^{p^s}\right)$ for some irreducible separable polynomial $g \in L[X]$. Each element $r_i^{p^s}$ (where the r_i are the roots of f) is a root of g and must be a simple root of g. Hence

$$g(X) = \left(X - r_i^{p^s}\right) g_i(X)$$

where $g_i(X) \in M[X]$ and $g_i\left(r_i^{p^s}\right) \neq 0$. Then

$$f(X) = (X - r_i)^{p^s} f_i(X)$$

where $f_i(r_i) = g_i\left(r_i^{p^s}\right) \neq 0$. This shows that $X - r_i$ is a factor of $f(X)$ of multiplicity exactly p^s. □

3.4 Some Galois Theory

In chapter 2 we used a theorem from Galois theory to prove the crucial Lemma 2.4.7. We will need that theorem here also so we will include a proof. The material of this section is standard. Our treatment follows [7].

Theorem 3.4.1. *Let G be a finite group of automorphisms of a field M and let L be the set of points of M left fixed by all of the elements of G. Then L is a subfield of M and $[M : L]$ equals the order of G. Moreover if σ is any automorphism of M fixing L, then σ in G.*

Proof. It is an easy exercise to see that L is a subfield of M. Let $|G| = n$ and $[M : L] = m$, and suppose $m < n$. Let $g_1, \ldots, g_n$ be the elements of G, and let $r_1, \ldots, r_m$ be a basis for M over L. Consider the following system of linear equations:

$$\sum_{1 \leq i \leq n} g_i(r_j)X_i = 0, \quad j = 1 \ldots m$$

Since there are n unknowns $X_1, \ldots, X_n$ and m equations, and we have assumed $m < n$, there is a nontrivial solution, say $x_1, \ldots, x_n$ in M. Then $\sum_{1 \leq i \leq n} x_i g_i$ is a L-linear map equal to 0 on every r_j, $1 \leq j \leq m$, hence on all of M:

$$\sum_{1 \leq i \leq n} x_i g_i = 0, \quad x_1, \ldots, x_n \neq 0\ . \tag{3.10}$$

The automorphisms $g_1, \ldots, g_n$ are therefore linearly dependent over M which is impossible according to lemma 3.4.2 below. Hence $m \geq n$.

Now assume $m > n$. Since $m > n$, there are $n + 1$ elements $r_1, r_2, \ldots, r_{n+1}$ of M which are linearly independent over L. Consider the following system of linear equations:

$$\sum_{1 \le i \le n+1} g_j(r_i) X_i = 0, \quad j = 1 \ldots n \tag{3.11}$$

Again we have more unknowns than equations so there is a nontrivial solution. This solution cannot lie in L because for some j, g_j is the identity and the corresponding equation would show that the $r_1, \ldots, r_{n+1}$ would be dependent over L.

Among all the nontrivial solutions pick one with the minimal number of elements different from zero. Rearranging the r_i, we can assume that exactly the first ℓ terms are not zero. Write this solution $s_1, s_2, \ldots, s_\ell, 0, 0, \ldots, 0$. By dividing by s_ℓ we may also assume that $s_\ell = 1$. Finally by permuting once more two r_i if necessary we can assume $s_1 \in M \setminus L$. Then

$$\sum_{1 \le i \le \ell-1} g_j(r_i) s_i + g_j(r_\ell) = 0, \quad j = 1 \ldots n \tag{3.12}$$

Pick some g_k with $g_k(s_1) \neq s_1$. Since G is a group, $g_k\, g_1$, $g_k\, g_2, \ldots, g_k\, g_n$ is a permutation of $g_1, g_2, \ldots, g_n$. Hence if we apply g_k to the system (3.12), we get

$$\sum_{1 \le i \le \ell-1} g_j(r_i) g_k(s_i) + g_j(r_\ell) = 0, \quad j = 1 \ldots n \tag{3.13}$$

If we subtract (3.13) from (3.12), we get

$$\sum_{1 \le i \le \ell-1} g_j(r_i)(s_i - g_k(s_i)) = 0, \quad j = 1 \ldots n \tag{3.14}$$

This shows that $(s_1 - g_k(s_1), s_2 - g_k(s_2), \ldots, s_{\ell-1} - g_k(s_{\ell-1}))$ is a nontrivial solution to (3.11) which has at most $\ell - 1$ elements different from zero: this contradicts the choice of ℓ.

If σ is another automorphism that stabilizes L, The size of the group generated by σ and G must also be $[M : L] = |G|$, therefore $\sigma \in G$. □

Lemma 3.4.2 (Dedekind Lemma)**.** *Let $g_1, \ldots, g_n$ be pairwise distinct automorphisms of some field E. Then they are linearly independent on E.*

Proof. Assume that $\sum_{i=1}^{n} x_i g_i = 0$ with $(x_1, \ldots x_n) \in E^{n*}$. Since any automorphism is nonzero, $n > 1$.

Removing elements in $g_1, \ldots, g_n$ if needed, assume that n is minimal. Pick $s \in M^*$ with $g_n(s) \neq g_1(s)$. For any r,

$$x_1\, g_1(s)\, g_1(r) + x_2\, g_2(s)\, g_2(r) + \cdots + x_n\, g_n(s)\, g_n(r) = 0. \qquad (3.15)$$

Hence $\sum x_i\, g_i(s) g_i = 0$; but one also has $\sum_{i=1}^{n} x_i g_n(s) g_i = 0$. Substracting these two equations yields a dependence relation between $g_1, \ldots, g_{n-1}$ which is nontrivial because the coefficient of g_1 is nonzero. This contradicts the minimality of n. □

3.5 The Norm Map

The results of this section are also standard and can be found in [19], for example.

Let M be a finite algebraic extension of a field L. Put $n = [M : L]$, the dimension of M as a vector space over L, and let $r_1, \ldots, r_n$ be a basis for M over L. For $r \in M$, let A_r be the matrix of the linear map on M given by multiplication by r with respect to the basis $r_1, \ldots, r_n$. Then the Cayley-Hamilton Theorem implies that $\det(r \cdot I \ - A_r) = 0$. Let $f_r(X)$ be the polynomial $\det(X \cdot I \ - A_r)$. We call f_r the *characteristic* polynomial of r with respect to the extension M of L. It is easy to see that f_r is independent of the choice of the basis $r_1, \ldots, r_n$. As remarked above, r satisfies f_r, however, f_r may not be irreducible. The minimal polynomial m_r is irreducible, and by Lemma 3.3.2 we see that $m_r | f_r$. Now f_r is clearly monic, *i.e.*, it can be written as

$$f_r(X) = X^n + c_1 X^{n-1} + \cdots + c_{n-1} X + c_n$$

with $c_i \in L$. In fact,

$$c_1 = -\operatorname{tr} A_r$$

and

$$c_n = (-1)^n \det A_r \ .$$

Definition 3.5.1. The *norm* of r is $\mathrm{N}(r) = (-1)^n c_n = \det A_r$, and the *trace* of r is $\mathrm{Tr}\ (r) = -c_1 = \operatorname{tr} A_r$.

Exercise 3.5.2. For $r, s \in M$ and $c \in K$, show the following:

(i) $\mathrm{N}(rs) = \mathrm{N}(r)\,\mathrm{N}(s)$.

(ii) If $r \in L$, then $\mathrm{N}(r) = r^n$.

(iii) $\operatorname{tr}(r + s) = \operatorname{tr}(r) + \operatorname{tr}(s)$.

(iv) $\operatorname{tr}(c \cdot r) = c. \operatorname{tr}(r)$.

(v) If $r \in L$, then $\mathrm{tr}(r) = n \cdot r$.

When it is necessary to indicate the fields involved, we write

$$\mathrm{N} = \mathrm{N}_{M/L}$$

and

$$\mathrm{tr} = \mathrm{tr}_{M/L}\,.$$

Now suppose Δ is a finite extension of M and let $r \in M$. We can consider r as a member of Δ and thus can consider $\mathrm{N}_{\Delta/L}(r)$ and $\mathrm{N}_{M/L}(r)$ and similarly for the two traces.

Proposition 3.5.3. *For $r \in M$, we have*

$$\mathrm{N}_{\Delta/L}(r) = [\mathrm{N}_{M/L}(r)]^m$$

and

$$\mathrm{tr}_{\Delta/L(r)} = m[\mathrm{tr}_{M/L}(r)]$$

where m is the degree of Δ over M.

Proof. Let $r_1, \ldots, r_n$ be a basis for M over L and $s_1, \ldots, s_m$ be a basis for Δ over M.

Exercise 3.5.4. Show that $\{r_i \cdot s_j : 1 \le i \le n \text{ and } 1 \le j \le m\}$ is a basis for Δ over L.

Order the elements of this basis by saying that $r_i \cdot s_j$ precedes $r_{i'} \cdot s_{j'}$ if $j < j'$ or if $j = j'$ if $i < i'$. Recall that A_r is the matrix of multiplication by r on M with respect to the basis $\{r_i\}$. Let C be the $mn \times mn$ matrix of multiplication by r on Δ with respect to the basis $\{r_i \cdot s_j\}$.

Exercise 3.5.5. Show that C is zero except for m blocks along the diagonal each containing the matrix A_r.

The lemma now follows easily. □

The matrix C in the above proof is the *m-fold tensor power* of A_r and we write $C = A_r^{(m)}$.

Corollary 3.5.6. *Let F_r (respectively f_r) be the characteristic polynomial of r with respect to the extension Δ (respectively M) of L. Then $F_r = f_r^m$.*

Proof. The proof follows trivially from the following equation:

$$F(X) = \det(X \cdot I \ - C) = \det(X \cdot I \ - A_r)^{(m)}. \tag{3.16}$$

□

This corollary allows us to prove the following interesting theorem which relates the minimal polynomial to the characteristic polynomial.

Theorem 3.5.7. *f_r is a power of m_r; $f_r = m_r$ if and only if $M = L(r)$.*

Proof. Let $\deg m_r = s$ and let f_1 be the characteristic polynomial of r considered as an element of $L(r)$. Now $[L(r) : L] = s$, so $\deg f_1 = s$ also. By Lemma 3.3.8, $m_r | f_1$, so $m_r = f_1$ which proves $f_r = m_r$ if $M = L(r)$. Now Corollary 3.5.6 above proves f_r is a power of m_r. Corollary 3.5.6 even shows that $f_r = [m_r]^m$ where $m = [M : L(r)]$, so if $f_r = m_r$ then $m = 1$ and $M = L(r)$. □

Proposition 3.5.8. *Let $L \subset M \subset \Delta$ be a tower of extensions as above. Let $r \in \Delta$. Then*

$$\mathrm{N}_{M(r)/L}(r) = [\mathrm{N}_{M/L} \circ \mathrm{N}_{M(r)/M}](r) \ .$$

Proof. By the above theorem, the minimal polynomial of r with respect to the extension $M(r)$ over M is the same as the characteristic polynomial. Suppose $\deg m_r = s$. Then it is clear that $\{1, r, r^2, \ldots, r^{s-1}\}$ is a basis for $M(r)$ over M. In fact, if we write

$$m_r(X) = X^s + c_1 X^{s-1} + \cdots + c_{s-1} X + c_s,$$

then

$$A_r = \begin{pmatrix} 0 & 1 & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \\ -c_s & -c_{s-1} & \cdot & \cdot & -c_2 & -c_1 \end{pmatrix}.$$

Now as in the proof of Proposition 3.3.10 we can get the matrix C of multiplication by r with respect to the extension $M(a)$ over L as follows: Suppose $[M : L] = n$. Let B_c be the matrix corresponding to multiplication by any element $c \in M$ for the extension M over L. So B_c is an $n \times n$ matrix, A_r is an $s \times s$ matrix, and C is an $ns \times ns$ matrix. To get C, replace each 0 (respectively 1) in A_r by the $n \times n$ zero (respectively identity) matrix, and replace c_i by the matrix B_{c_i}.

Expanding the determinant of A_r, we get

$$\mathrm{N}_{M(r)/L}(r) = \det B_{(-1)^s c_s} \ .$$

Since

$$c_s = (-1)^n \, \mathrm{N}_{M(r)/M}(r),$$

the proposition follows. □

Corollary 3.5.9. *Let M be a finite algebraic extension of L and Δ a finite algebraic extension of M. Then*

$$\mathrm{N}_{\Delta/L} = \mathrm{N}_{M/L} \circ \mathrm{N}_{\Delta/M} \quad .$$

The other roots of the minimal polynomial, m_r, of r are called the *conjugates* of r. They may or may not be in M. If M contains the conjugates of each of its elements, we say M is a *normal* extension of L.

Elements that are conjugates are closely related as the next theorem will show. First we need a lemma.

Lemma 3.5.10. *Let $\sigma : L \to L'$ be an isomorphism between two fields. Let $f(X) \in L[X]$ be irreducible, and set $g = \sigma(f)$, i.e., apply σ to the coefficients of f. Let r (respectively s) be a root of f (respectively g) in some extension field. Then σ can be extended to an isomorphism from $L(r)$ to $L'(s)$.*

Proof. Since r is algebraic over L, every element in $L(r)$ is of the form $h(r)$ for some $h(X) \in L[X]$. Furthermore if we take h to have degree $\leq n-1$ where $n = \deg f$, then h is unique (since $f = m_r$). Now we map $h(r) \in L(r)$ into $[\sigma(h)](s) \in L'(s)$, and it is not hard to see that this is an isomorphism. □

It should be clear from this lemma that if M is algebraic over L, and r and r' are conjugate over L, then there is an isomorphism: $L(r) \to L(r')$ taking r into r' and fixing L. We can, however, do better if M is normal over L. If M is normal over L, we can extend this isomorphism to all of M.

Theorem 3.5.11. *If M is a finite normal extension of L and $r, r' \in M$ are conjugate over L, then there is a automorphism of M taking r into r' which leaves L fixed.*

Proof. Let $r_1, r_2, \ldots, r_m \in M$ generate M over L. Let

$$F(X) = \prod_{i=1}^{m} m_{r_i}(X) \in L[X] \ .$$

F is monic, and its roots generate M. Since M is normal over $L, F(X)$ factors completely in $M[X]$. Let $\deg F = n$. We are going to prove the following somewhat more general result: Suppose we are given an isomorphism $\sigma\colon\ L_1 \to L_2$ between two fields. Further suppose we have monic polynomials $F_1(X)$ and $F_2(X)$ in $L_1[X]$ and $L_2[X]$ respectively with $\sigma(F_1) = F_2$. Further suppose for $i = 1$, 2 there are fields M_i which are generated by the roots of F_i and in which F_i factors completely *i.e.*, into linear factors. Then τ can be extended to an isomorphism from M_1 to M_2.

First we show our result follows from the more general result. Let $\sigma : L(r) \to L(r')$ be the isomorphism given by the lemma using m_r and $m_{r'}$ as the irreducible polynomials. Hence $\sigma(r) = r'$ and σ leaves L fixed. Since $F(X) \in L[X], \sigma(F) = F$. Now we are in the situation of our more general result with $L_1 = L(r)$, $L_2 = L(r')$, and $F_1 = F_2 = F$.

The proof of the more general result is by induction on n. The case $n = 1$ is trivial. For the inductive step, let G_1 be a factor of F_1 which is irreducible over L_1. Then $G_2 = \sigma(G_1)$ will be a factor of F_2 irreducible over L_2. Pick a root $s_1 \in M$ of G_1 and a root

$s_2 \in M$ of G_2. Now apply the lemma again using G_1, G_2, s_1, and s_2 to get an extension of σ to $\tau : L_1(s_1) \to L_2(s_2)$ with $\tau(s_1) = s_2$. We can write $F_1(X) = (X - s_1)F_3(X)$ and $F_2(X) = (X - s_2)F_4(X)$. Now $F_3(X) \in L_1(s_1)[X]$ and $F_4[X] \in L_2(s_2)[X]$, and since $\tau(s_1) = s_2, \tau(F_3) = F_4$. Also F_3 and F_4 are monic and of degree $n-1$. Now F_3 factors completely in M_1, and M_1 is generated by the roots of F_3 over $L_1(s_1)$, and F_4 factors completely in M_2 and M_2 is generated by the roots of F_4 over $L_2(s_2)$. Hence by induction, we get an extension of τ to an isomorphism $p : M_1 \to M_2$. Since τ is an extension of σ, we have proved the more general result.

Since in our case, $\sigma(r) = r'$ and σ leaves L fixed, we are done.

□

There is at least one case where the extension is normal and separable, namely the situation of Theorem 3.4.1. This result and its converse are proven below. An extension which is both normal and separable is said to be *Galois*.

Theorem 3.5.12. *M is a Galois extension of L if and only if L is the fixed field of a group of automorphisms of M.*

Proof. Assume L is the fixed field of a group G of automorphisms of M. Let $g_1, g_2, \ldots, g_n$ be the elements of the group G. Let r be an arbitrary element of M. We will show that the minimal polynomial of r has distinct roots all of which are in M. Let $r, r_2, r_3, \ldots, r_m$ be the set of distinct elements in the sequence $g_1(r), g_2(r), \ldots, g_n(r)$. Since G is a group of automorphisms, the elements $r, r_2, r_3, \ldots, r_m$ are permuted by the elements of G. Consider the polynomial

$$f(X) = (X - r)(X - r_2) \cdots (X - r_m) \ .$$

Since an automorphism in G merely permutes the factors of f, the coefficients of f are left fixed by all of the automorphisms in G. Hence $f(X) \in L(X)$. Clearly $f(r) = 0$ so if we can show that f has minimum degree among those polynomials in $L[X]$ with r as a root then f will be the minimal polynomial of r, and we will be half done. Suppose g is any polynomial in $L[X]$ with r as a root. Applying an automorphism in G to the equation $g(r) = 0$, we get $g(r_i) = 0$ for some i, in fact, for each i. Thus the degree of g must be at least m, the degree of f, and we are half done.

Now assume M is Galois over L and let $g_1, g_2, \ldots, g_n$ be the automorphisms of M which fix L. It is easy to see that the g_i form a group. If $r \in M$ is left fixed by all of the g_i, then by Theorem 3.5.11, r can have no conjugates in M and hence since M is separable over L, r must be in L. Therefore L is the fixed field of the g_i. □

In this situation, the group of automorphisms of M that fix L is called the *Galois* group of M (or M over L).

Let us assume $M = L(r)$ and that M is normal. Then we can write

$$m_r(X) = \prod_{i=1}^{m} (X - r_i)$$

where $r_1 = r$ and the other r_i's are the conjugates of r and are in M.

Theorem 3.5.7 tells us that $m_r = f_r$ in this case. Therefore we get

$$\mathrm{N}(r) = \prod_{i=1}^{n} r_i \qquad (3.17)$$

and

$$\mathrm{tr}(r) = \sum_{i=1}^{n} r_i.$$

If r is separable over L, then the $r_1, \ldots, r_n$ are distinct by Corollary 3.3.12. If r is inseparable over L, and if n_0 is the degree of separability of f_r, and p^s is its degree of inseparability, then by Proposition 3.3.13,

$$f_r(X) = \prod_{i=1}^{n_0} (x - r_i)^{p^s} \ ,$$

so r has n_0 distinct conjugates and $n_0 p^s = \deg f_r$. Hence the above two equations yield

$$\mathrm{N}(r) = \left(\prod_{i=1}^{n_0} r_i \right)^{p^s} \tag{3.18}$$

and

$$\mathrm{tr}(r) = p^s \cdot \left(\sum_{i=1}^{n_0} r_1 \right) = 0.$$

We restate the last equation in a proposition.

Proposition 3.5.13. *If M is a finite extension of L and $r \in M$ is inseparable over L, then $\mathrm{tr}_{M/L}(r) = 0$.*

3.6 The Norm Map (Continued)

Now suppose we are given an isogeny $\alpha : E \to E'$ between elliptic curves E and E', and we put $K' = \alpha^*(K(E')) \subset K(E)$. We would like to compute the norm $\mathrm{N} = \mathrm{N}_{K(E)/K'}$. We will show that

$$\forall\, r \in K(E), P \in E, \quad \mathrm{N}(r)(P) = \left(\prod_{\alpha(Q)=\alpha(P)} r(Q) \right)^{e_\alpha} \tag{3.19}$$

where e_α is the ramification index of α.

Now this is very close to Equation (3.18) which says that $\mathrm{N}(r)$ is the product of the conjugates of r over K'. We have to identify the conjugates of the rational function r with the translates of r by points in E that get sent into $\mathcal{O}'$ by α. We also must show that the ramification index of α is p^s, the degree of inseparability of f_r, the characteristic polynomial of r over K'. Incidentally, this will also clear up a loose end from Section 3.3 where we wanted to know how the separability of α was reflected in the extension of K'. Obviously the answer must be that the extension is separable, but we have not proved it as yet.

Let $\mathcal{S} = \ker \alpha \subset E$, and consider the group of translations

$$\mathcal{T}_\mathcal{S} = \{\mathcal{T}_P : P \in \mathcal{S}\} \ .$$

Then $\mathcal{T}_\mathcal{S}$ is a finite group of automorphisms of E. $\mathcal{T}_\mathcal{S}$ also acts on $K(E)$ by

$$\mathcal{T}_P(r) = \mathcal{T}_P^*(r) = r \circ \mathcal{T}_P \ .$$

Now suppose $r \in K'$, so $r = r' \circ \alpha$ for some $r' \in K[E']$. Then

$$[\mathcal{T}_P(r)](Q) = r'(\alpha(Q + P)) = r'(\alpha(Q)) = r(Q) \ ,$$

so $\mathcal{T}_\mathcal{S}$ leaves K' fixed. Set

$$L = \{r \in K(E) : \mathcal{T}_P(r) = r, \quad \forall P \in \mathcal{S}\} \ .$$

Then Theorem 3.4.1 tells us that $K(E)$ is a finite algebraic extension of L of degree $m = |\mathcal{S}|$, and Theorem 3.5.12 tells us that this extension is Galois.

The idea of our proof of (3.19) is that N is the composition of the norm from $K(E)$ down to L with the norm from L down to K'.

We will show that there is an elliptic curve C with $K(C) = L$ such that the map α factors through C. The point is that since $K(E)$ is Galois over L, the factor from E to C should be "nice", while the factor from C to E' should be very "special".

Let

$$\tilde{x} = [\mathrm{tr}_{K(E)/L}](x)$$

$$\tilde{y} = [\mathrm{tr}_{K(E)/L}](y),$$

so

$$\tilde{x} = \sum_{P \in \mathcal{S}} \mathcal{T}_P(x)$$

$$\tilde{y} = \sum_{P \in \mathcal{S}} \mathcal{T}_P(y)\ .$$

Note that each $\mathcal{T}_P(x)$ (resp. $\mathcal{T}_P(y)$) has a pole of degree 2 (resp. 3) at $-P$; therefore $\tilde{x}$ (resp. $\tilde{y}$) is nonzero and has a pole of multiplicity 2 (respectively 3) at each point of $\mathcal{S}$ and no other poles.

Proposition 3.6.1. *$\tilde{x}$ and $\tilde{y}$ generate L.*

Proof. Clearly $\tilde{x}$ and $\tilde{y}$ are in L. Let $r \in L$, and suppose r is even, *i.e.*, $r(-P) = r(P)$. Substituting r by $1/r$ if necessary, further assume that $\mathrm{ord}_{\mathcal{O}}(r) \leq 0$. If we multiply r by a suitable power of

$$\prod_{\substack{P \notin \mathcal{S} \\ r(P) = \infty}} (\tilde{x} - \tilde{x}(P)),$$

which is nonzero, we will get an even rational function r_1 which is in L and whose poles are only on $\mathcal{S}$.

Now r_1 even implies that $\mathrm{ord}_{\mathcal{O}}(r_1)$ is even. Say $\mathrm{ord}_{\mathcal{O}}(r_1) = -2a \leq 0$. If we subtract a suitable scalar multiple of $\tilde{x}^a$ from r_1, we will get an even function r_2 in L whose only poles are on $\mathcal{S}$ and s.t. $\operatorname{ord}_{\mathcal{O}}(r_2) \geq -2a+2$. We continue this process until $\operatorname{ord}_{\mathcal{O}}(r_k) > 0$ (which can be achieved by adding only positive powers of $\tilde{x}$). Then

$$r_k = r_1 + p(\tilde{x})$$

where p is a polynomial, and r_k is in L with no poles off $\mathcal{S}$ and with $r_k(\mathcal{O}) = 0$. But $r_k \in L$ implies $r_k(P) = r_k(\mathcal{O})$ for any $P \in \mathcal{S}$. Thus r_k has no poles at all, so r_k is zero, and r_1 and r are in $K(\tilde{x})$.

Now suppose $r \in L$ is odd. Then $\tilde{y}r$ is even, so $\tilde{y}r \in K(\tilde{x})$ and $r \in K(\tilde{x}, \tilde{y})$. Since every rational function is the sum of an even function and an odd function, we are done. □

Theorem 3.6.2. *Let L the field fixed by the group of translations $\{\mathcal{T}_P \,|\, P \in \ker\alpha\}$. Then $K(E) : L$ is Galois and there is an elliptic curve C such that*

(i) $K(C)$ is isomorphic to L

(ii) there are homomorphisms $\beta : E \to C$ and $\gamma : C \to E'$ such that

(a) $\alpha = \gamma \circ \beta$,

(b) $\ker\beta = \ker\alpha$,

(c) $e_\beta = 1$,

(d) $\ker\gamma = \left\{\tilde{\mathcal{O}}\right\}$ where $\tilde{\mathcal{O}}$ is the identity element of C, and

(e) $e_\gamma = e_\alpha$.

Proof. $\tilde{y}^2$ is even so $\tilde{y}^2 \in K(\tilde{x})$. Furthermore $\operatorname{ord}_{\mathcal{O}}(\tilde{y}^2) = 6$, and $\tilde{y}^2$ has no pole outside $\mathcal{S}$. Therefore reasoning as for the proof of proposition 3.6.1, one sees that

$$\tilde{y}^2 = f(\tilde{x}) \tag{3.20}$$

for some cubic polynomial f. If we can show that the polynomial f has three distinct roots, then we can let C be the elliptic curve defined by (3.20). Then part i) will follow from Proposition 3.6.1 above. Remark that

- $\tilde{x}$ is even, $\tilde{y}$ is odd. For instance, $\tilde{y}(-P) = \sum_{Q \in \mathcal{S}} y(Q - P) = \sum_{Q \in \mathcal{S}} y(-Q - P) = -\tilde{y}(P)$.

- If $P - Q$ is in $\mathcal{S}$, $\tilde{x}(P) = \tilde{x}(Q)$ and $\tilde{y}(P) = \tilde{y}(Q)$, because $\mathcal{S} + P = \mathcal{S} + Q$.

- conversely, assume $\tilde{x}(P) = \tilde{x}(Q)$ and $\tilde{y}(P) = \tilde{y}(Q)$. Then by proposition 3.6.1, for every $r \in L$, $r(P) = r(Q)$ *i.e.* $\mathcal{T}_{P-Q}$ is the identity on L. By theorem 3.4.1 $\mathcal{T}_{P-Q}$ must be in the Galois group of $K(E) : L$ which is $\ker \alpha$ by construction: $P - Q \in \mathcal{S}$.

Suppose $f(k) = 0$ for some $k \in K$. Let $N = |\{P \in E : \tilde{x}(P) = k\}|$. By (3.20), $\tilde{x}(P) = k$ implies $\tilde{y}(P) = 0$. Hence if $\tilde{x}$ takes the value k at both P and Q in E, then both $\tilde{x}$ and $\tilde{y}$ take the same values at P and Q, and $\alpha(P) = \alpha(Q)$: α takes the same value at each of the points with $\tilde{x}(P) = k$, *i.e.* α maps N points of E to one point of E', which implies

$$N \leq |\ker \alpha| \ .$$

On the other hand, $\tilde{x}$ is invariant under $\mathcal{T}_\mathcal{S}$ so $\tilde{x}$ takes the value k at least $|\mathcal{S}|$ times: $N \geq |\ker\alpha|$ and finally

$$N = |\ker\alpha|. \tag{3.21}$$

Consider the set

$$\mathcal{R} = \{P \in E : \tilde{y}(P) = 0\} .$$

If $\tilde{y}$ is zero at P, then $\tilde{x}(P) = k$ for some root k of f. If we let M be the number of roots of f, then (3.21) tells us that $|\mathcal{R}| = M \cdot |\ker\alpha|$, *i.e.*, $\tilde{y}$ is zero at precisely $M \cdot |\ker\alpha|$ points of E.

Suppose $P \in E$ is such that $2\,\alpha(P) = \mathcal{O}'$, but $\alpha(P) \neq \mathcal{O}'$. Since $P + P$ is in $\mathcal{S}$, $\tilde{y}(P) = \tilde{y}(-P)$; but since $\tilde{y}(-P) = -\tilde{y}(P)$, $\tilde{y}(P) = 0$ (again, the characteristic of K is not 2 or 3).

We know there are three points in $E'[2] \setminus \{\mathcal{O}\}$. Since α is surjective, α maps $|\ker\alpha|$ points into each one. So by the above, $\tilde{y}$ is zero at at least $3 \cdot |\ker\alpha|$ points. Hence f has at least 3 roots, and since f is cubic, it must have distinct roots, and i) follows.

Since $\tilde{x}(P) = \tilde{x}(Q)$ and $\tilde{y}(P) = \tilde{y}(Q)$ iff $P - Q \in \mathcal{S}$, C is the quotient group $E/\mathcal{S}$. We take $\beta : E \to C$ to be the canonical projection

$$\beta(P) = (\tilde{x}(P), \tilde{y}(P)) .$$

Then γ is the right factorization of α by β. a), b), d) follow.

We have considered $\tilde{x}$, and $\tilde{y}$ as functions on E. Of course, they can also be considered as functions on C. In fact, they are the coordinate functions on C. With this “abuse of terminology” we see that $\tilde{x} \circ \beta = \tilde{x}$ where the first $\tilde{x}$ is a function on C, and the second is a function on E. Seen either as a function on C or

E, $\tilde{x}$ satisfies $\mathrm{ord}_{\mathcal{O}}\, \tilde{x} = -2$ (otherwise this abuse of notation would put us in trouble). As a consequence $e_\beta = 1$ and c) follows. Since $e_\alpha = e_\beta \cdot e_\gamma$, we get e). □

We have now split up the map α into a map β that has a kernel but no ramification and a map γ that has ramification but no kernel. One of the results we want from all this is that if α is separable (*i.e.*, $e_\alpha = 1$), then $K(E)$ is a separable extension of K'.

Proposition 3.6.3. *Let $\alpha : E \to E'$ be an injective homomorphism. Then α is an isomorphism if and only if $e_\alpha = 1$ or $K(E)$ is separable over K'.*

Proof. Let $\overline{x} = \alpha^*(x')$, and $\overline{y} = \alpha^*(y')$ where x' and y' are the coordinates on E'. Recall $K' = \alpha^*(K(E')) \subset K(E) = K(x, y)$. Since $\overline{x}$ is an even function on $E, \overline{x} \in K(x)$.

Claim. If we consider $\overline{x}$ as a function of x, it takes each value of K exactly once. Suppose $k, l \in K$ and $\overline{x}(k) = \overline{x}(l)$. Let $P, Q \in E$ with $x(P) = k$ and $x(Q) = l$. Then

$$\alpha(P) = (\overline{x}(P), \overline{y}(P)) = (\overline{x}(Q), \pm\overline{y}(Q)) = \pm\alpha(Q)\ .$$

By replacing Q with $-Q$ we can assume that $\alpha(P) = \alpha(Q)$. Since α is injective by hypothesis and surjective by Proposition 3.1.2, the claim follows.

Since $\alpha(P) = \mathcal{O}'$ implies $P = \mathcal{O}$, we see that $\overline{x}$ has no finite poles. Hence $\overline{x}$ is a polynomial. Exercise 3.6.4 below shows that $\overline{x}$ can be written $(ax + b)^{p^r}$.

Now x' has a pole of multiplicity 2 at $\mathcal{O}' \in E'$ so $\overline{x}$ has a pole of multiplicity $2e_\alpha$ at $\mathcal{O}$. The above equation shows that we must have $e_\alpha = p^r$.

Since $\overline{x} = \alpha^*(x')$, if $\overline{x}$ had a p^{th}-root in K', x' would have a p^{th}-root in $K(E')$ which is not the case. Thus by Lemma 3.3.8 the minimal polynomial of $ax + b$ over K' is $X^{p^r} - \overline{x}$, and must be irreducible. If $K(E)$ is separable over K' this forces $r = 0$. Hence $K(E)$ separable implies $e_\alpha = 1$.

Let us now show that if $e_\alpha = 1$ then $K' = K(E)$ and α is an isomorphism. If $e_\alpha = 1$, then $\overline{x} = ax + b$ so $x \in K'$. Now $\overline{y}$ is an odd function, and since $\alpha^{-1}(\mathcal{O}') = \{\mathcal{O}\}$, $\overline{y}$ also has no finite poles. Hence $\overline{y} = yh$ with $h \in K[x]$, and y is also in K', therefore $K' = K(E)$. Since $e_\alpha = 1$, $\operatorname{ord}_{\mathcal{O}}(\overline{x}) = -2$. Since $\overline{y}^2 = \overline{x}^3 + A'\overline{x} + B'$, $\operatorname{ord}_{\mathcal{O}}(\overline{y}) = -3$ and $\operatorname{ord}_{\mathcal{O}}(h) = 0$. This implies that h is a nonzero constant. Therefore $e_\alpha = 1$ yields

$$\overline{x} = ax + b \text{ and } \overline{y} = cy$$

for some $a, b, c \in K$ and since $\alpha(P) = (\overline{x}(P), \overline{y}(P))$, α is manifestly an isomorphism.

Conversely, if α is an ismorphism, $e_\alpha = 1$. We have seen that this implies that $K(E) = K'$, and therefore $K(E)$ is separable over K'.

□

Exercise 3.6.4. Let K be a field of characteristic $p > 0$ and φ a polynomial of $K[x]$ taking each value of K exactly once. Then φ can be written

$$\varphi(x) = (ax + b)^{p^r} \tag{3.22}$$

for some $a, b \in K$ and $r \geq 0$.

Proof. Let r be s.t. $\varphi = \psi^{p^r}$ and $\psi' \neq 0$. Then ψ must be injective. Let $d = \deg \psi$. Since K is algebraically closed, if $d > 0$, $\psi - u$ has d roots counted with multiplicity for any $u \in K$. Hence $d = 1$. □

Corollary 3.6.5. *Let $\alpha : E \to E'$ be a homomorphism. Then α is separable (i.e., $e_\alpha = 1$) if and only if $K(E)$ is a separable extension of K'.*

Proof. We use theorem 3.6.2 to reduce the general case to the case where α is injective. Write $\alpha = \gamma \circ \beta$ as in Theorem 3.6.2. Then α is separable if and only if γ is, and γ is injective. Theorem 3.6.2 also says that $K(E)$ is Galois, hence separable, over $K(C)$, therefore $K(E)$ is separable over K' if and only if $K(C)$ is separable over K', *i.e.* iff γ is separable by proposition 3.6.3. □

Corollary 3.6.6. *Let $\alpha : E \to E'$ be a homomorphism. If α has trivial kernel, then after a suitable change of coordinates, we get $\alpha(x, y) = (x^{p^r}, y^{p^r})$, and $e_\alpha = p^r$ for some $r > 0$. K' is then equal to $K(E)^{p^r}$.*

Proof. This is really a corollary of the proof of proposition 3.6.3. Equation (3.22) tells us that after composing with an isomorphism, $\overline{x} = x' \circ \alpha = x^{p^r}$. Since $\overline{y}$ is an odd function, we can write $\overline{y} = yh$ where h is a function of x alone. It is easily seen that h has no finite poles and thus is a polynomial. Since $\overline{y}^2 = \overline{x}^3 + A'\overline{x} + B'$, we get

$$\begin{aligned} y^2h^2 &= x^{3p^r} + A'x^{p^r} + B' \\ &= s(x)^{p^r} + \left(A' - A^{p^r}\right) x^{p^r} + \left(B' - B^{p^r}\right) \end{aligned}$$

where $s(x) = x^3 + Ax + B$. Now $y^2 = s(x)$ so s divides $C^{p^r}x^{p^r} + D^{p^r} = (Cx + D)^{p^r}$ where we have put $A' - A^{p^r} = C^{p^r}$ and $B' - B^{p^r} = D^{p^r}$ (this is possible since the base field is algebraically closed). Hence s must divide $Cx + D$ for $C, D \in K$ which is absurd.

Therefore we must have $C = D = 0$, and it follows that $\overline{y} = y^{p^r}$. The part about the ramification index follows from the definition.

□

To summarize what we have proved in this section so far, we see that a homomorphism between elliptic curves can be factored into a separable part and a factor that looks like $\left(x^{p^r}, y^{p^r}\right)$ (such maps are called *purely inseparable*). We will use this result to prove the formula for the norm by breaking it up into two easier cases.

Theorem 3.6.7. *Let $\alpha : E \to E'$ be a homomorphism, and $K' = \alpha^*(K(E')) \subset K(E)$. Let $\mathrm{N} = \mathrm{N}_{K(E)/K'}$, and $r \in K(E)$. Then*

$$\forall\, P \in E\,, \quad [\mathrm{N}(r)](P) = \prod_{\alpha(Q)=\alpha(P)} r(Q)^{e_\alpha}$$

Proof. This theorem is now a consequence of the previous results. Theorem 3.6.2 says that $\alpha = \gamma \circ \beta$ where β is separable and γ is purely inseparable. Corollary 3.5.9 tells us that the corresponding norm is also a composition of the norms on $K(E) : \beta^*(K(C))$ (separable case) and $\beta^*(K(C)) : \beta^*(\gamma^*(K(E')))$ (inseparable case). The latter is isomorphic to $K(C) : \gamma^*(K(E'))$ since a nonzero field homomorphism is always an isomorphism. In the separable case Theorem 3.5.12 together with the proof of Theorem 3.6.2 says that the Galois group of $K(E)$ over $\beta^*(K(C))$ is $\mathcal{T}_S(\approx \ker\beta)$ so the conjugates (over $\beta^*(K(C))$) of a rational function r are the functions $\mathcal{T}_P(r)$ for $P \in \ker\beta$. In the purely inseparable case by Corollary 3.6.6, $\gamma = \left(x^{p^r}, y^{p^r}\right)$ for some $r > 0$, and $e_\alpha = p^r$. It remains to prove that the degree of inseparability of any v in $K(E)$ over $\gamma^*(K(E))$ is also p^r. $X^{p^r} - v^{p^r}$ cancels v and by reasoning as in the proof of lemma 3.3.8, one sees that the minimal polynomial of v is

$(X - v)^{p^{r_v}} = X^{p^{r_v}} - v^{p^{r_v}}$ with $r_v \leq r$ the smallest integer s.t. $v^{p^{r_v}}$ is in $\gamma^*(K(E)) = K(E)^{p^r}$. As a consequence, if the degree of the extension $K(E) : \gamma^*(K(E))$ is p^r, the characteristic polynomial of v on $\gamma^*(K(E))$ is $X^{p^r} - v^{p^r}$ and the inseparability degree of v is p^r as contended. But x is of degree p^r over $K(E)^{p^r}$ (since $K(E)$ contains a copy of $K[x]$); let us show that $K(E)^{p^r}[x] = K(E)$. $K(E)^{p^r}[x]$ contains y^2 and y^{p^r}; since p is odd ($p \neq 2$ and $p \neq 3$), $K(E)^{p^r}[x]$ contains $y^{p^s \mod 2} = y$ and is indeed equal to $K(E)$. Our result now follows.

□

There are a number of important consequences of this theorem.

Definition 3.6.8. Let $\alpha : E \to E'$ be an isogeny. Define

$$\alpha_* : \mathrm{div}(E) \to \mathrm{div}(E')$$

by setting

$$\alpha_*(\langle P \rangle) = \langle \alpha_*(P) \rangle$$

and extending linearly. We also define $\alpha_* : K(E) \to K(E')$ by

$$\alpha_*(r) = N(r) \circ \alpha^{-1}$$

where $r \in K(E)$, and

$$N = N_{K(E)/\alpha^*(K(E'))} \ ,$$

i.e., $N(r) = r' \circ \alpha$ for some $r' \in K(E')$, and $\alpha_*(r) = r'$.

Theorem 3.6.9 ("Lower Star"). *Let $\alpha : E \to E'$ be an isogeny. For $r \in K(E)$,*

$$\alpha_*(\mathrm{div}\, r) = \mathrm{div}(\alpha_*(r)) \ .$$

Proof. It follows immediately from Theorem 3.6.7 that

$$[\alpha_*(r)](P') = \prod_{\alpha(Q)=P'} r(Q)^{e_\alpha}$$

for $P' \in E'$. While the theorem follows by simply unraveling this equation, it is perhaps easier to understand if we do our usual trick of breaking it up into the separable and purely inseparable cases.

In the separable case ($e_\alpha = 1$), we get

$$\begin{aligned}\operatorname{div} \alpha_*(r) &= \sum_{P \in E'} \sum_{\alpha(Q)=P'} \operatorname{ord}_Q(r) \, \langle P' \rangle \\ &= \sum_{Q \in E} \operatorname{ord}_Q(r) \, \langle \alpha_*(Q) \rangle \\ &= \alpha_*(\operatorname{div}(r)) \ .\end{aligned}$$

Now in the purely inseparable case ($\ker \alpha = \mathcal{O}$), we can assume $\alpha(x, y) = (x^{e_a}, y^{e_a})$, so

$$[\alpha_*(r)](P') = r(P)^{e_a}$$

where $P \in E$ is the unique point with $\alpha(P) = P'$. Hence

$$[\alpha_*(r) \circ \alpha](P) = r(P)^{e_\alpha}$$

so

$$\operatorname{ord}_P(\alpha_*(r) \circ \alpha) = e_\alpha \cdot \operatorname{ord}_P(r) \ .$$

On the other hand, by Proposition 3.1.4,

$$\operatorname{ord}_P(\alpha_*(r) \circ \alpha) = [\operatorname{ord}_{P'} \alpha_*(r)] e_\alpha \ ,$$

and we see that

$$\mathrm{ord}_{P'}\ (\alpha_*(r)) = \mathrm{ord}_P(r)\ .$$

Finally

$$\begin{aligned}\mathrm{div}(\alpha_*(r)) &= \sum_{P \in E'} \mathrm{ord}_{P'}(\alpha_*(r))\,\langle P' \rangle \\ &= \sum_{P \in E} \mathrm{ord}_P(r)\,\langle \alpha_*(P) \rangle \\ &= \alpha_*(\mathrm{div}(r))\ .\end{aligned}$$

□

Exercise 3.6.10.

(i) Show $\alpha_* \circ \alpha^*$ acts as multiplication by $\deg \alpha$ on $\mathrm{div}(E')$.

(ii) If $\beta : E' \to E'$ is another rational mapping, then

$$(\beta \circ \alpha)_* = \beta_* \circ \alpha_*.$$

3.7 Weil Reciprocity

Let $D = \Sigma n_P \langle P \rangle$ be a divisor on E. We define the *support* of D to be the set of points $P \in E$ such that $n_p \neq 0$.

Definition 3.7.1. Let r be a rational function on E and D a divisor on E, and suppose $\mathrm{div}\, r$ and $D = \Sigma n_P \langle P \rangle$ have disjoint supports. Then we can define

$$r(D) = \prod_{P \in E} f(P)^{n_P}\ .$$

Exercise 3.7.2. Let $\alpha : E \to E'$ be a rational map. Prove the following two equations in the sense that if one side is well-defined, then so is the other and they are equal:

(i) $r(\alpha^*(D')) = [\alpha_*(r)](D')$ for $r \in K(E)$ and $D' \in \operatorname{div}(E')$.

(ii) $r'(\alpha_*(D)) = [\alpha^*(r')](D)$ for $r' \in K(E')$ and $D \in \operatorname{div}(E)$.

Now Weil reciprocity can be stated very simply. Suppose r and s are rational functions on E whose divisors have disjoint supports. Then

$$r(\operatorname{div} s) = s(\operatorname{div} r). \tag{3.23}$$

Unfortunately we have not been able to find a simple proof. r and s cannot both be polynomials since all polynomials have a pole at $\mathcal{O}$, but if they were monic polynomials, and if they were both functions of one variable, then (3.23) would say that the product of the values of r on the zeros of s equals the product of the values of s on the zeros of r (up to sign). This is well-known to be true. It is merely the fact that the resultant of two polynomials is symmetric. Weil reciprocity can be thought of as the generalization of this to three rational functions of two variables, namely r, s, and the polynomial $y^2 - x^3 - Ax - B$ which defines the elliptic curve E.

The usual method of proof of this result is to first prove it on the projective line $\mathbb{P}_1$ where it is easy, and then to use the "Lower Star" theorem to pull the result back to the elliptic curve considering one of our rational functions as a rational mapping between E and $\mathbb{P}_1$. For us, this approach has the difficulty that $\mathbb{P}_1$ is not an elliptic curve so our proof of the "Lower Star" Theorem is not valid. The difficulty is that rational maps to $\mathbb{P}_1$ are not all homomorphisms. One might think that since $\mathbb{P}_1$ is such a simple object, the proof should be easier in this case, but we have not been able to find any proof except the very general one which works for maps between any two smooth curves.

We have decided to adopt a slightly different approach. There is a generalization of Weil reciprocity that essentially removes the requirement that $\operatorname{div} f$ and $\operatorname{div} g$ have disjoint supports. We will state this theorem and then do enough valuation theory to state the main result from the valuation theory we need. We will sketch the proof of the valuation theoretic result in an Appendix. A reference is [16], Chapter III, Section 1, especially page 45. We would like to thank Noam Elkies for pointing out this generalization.

We begin with a definition. We assume $f, g \neq 0$ in what follows.

Definition 3.7.3. Let E be an elliptic curve and $f, g \in K(E)$. For $P \in E$ we set

$$\langle f, g\rangle_P = (-1)^{mn}\left[\frac{f^n}{g^m}\right](P) \quad , \quad m = \operatorname{ord}_P f \text{ and } n = \operatorname{ord}_P g.$$

We call $\langle f, g\rangle_P$ the *local symbol* of f and g. The local symbol is sometimes called the *tame symbol.*

The following exercises are all quite easy.

Exercise 3.7.4.

(i) Let $h = \dfrac{f^n}{g^m}$. Show that $\operatorname{ord}_P h = 0$ so that the above definition makes sense.

(ii) $\langle f, g\rangle_P = 1$ unless f or g has a pole or zero at P.

(iii) If $1 - g$ has a zero at P then $\langle f, g\rangle_P = 1$.

(iv) Suppose $\operatorname{ord}_P f = 0$, then $\langle f, g\rangle_P = f(P)^n$.

(v) $\langle f, hg\rangle_P = \langle f, g\rangle_P \cdot \langle f, h\rangle_P$, and $\langle fh, g\rangle_P = \langle f, g\rangle_P \cdot \langle h, g\rangle_P$ for any $h \neq 0$.

(vi) $\langle f, g\rangle_P \cdot \langle g, f\rangle_P = 1$.

(vii) $\langle -f, f\rangle_P = 1$.

(viii) $\langle 1 - f, f\rangle_P = 1$.

Proof.

(i) clear.

(ii) clear.

(iii) $\operatorname{ord}_P g = 0$; with $m = \operatorname{ord}_P(f)$, $\langle f, g\rangle_P = \frac{f^0}{g^m}(P) = 1$.

(iv) Direct consequence of the definition.

(v) For instance for the right linearity: $\operatorname{ord}_P hg = \operatorname{ord}_P h + \operatorname{ord}_P g$ hence

$$\begin{aligned}\langle f, hg\rangle_P &= (-1)^{\operatorname{ord}_P(f)\cdot(\operatorname{ord}_P h + \operatorname{ord}_P g)} \left[\frac{f^{(\operatorname{ord}_P h + \operatorname{ord}_P g)}}{(gh)^{\operatorname{ord}_P f}}\right](P) \\ &= \langle f, g\rangle_P \cdot \langle f, h\rangle_P\end{aligned}$$

(vi) Direct consequence of the definition.

(vii) Direct consequence of the definition. Note also that $\langle f, f\rangle_P = (-1)^{\operatorname{ord}_P f}$.

(viii) If neither f nor $1-f$ have a pole or a zero at P, $\langle f, 1-f\rangle_P = 1$. If one of f and $1-f$ has nozero order at P, one can assume it is f by symmetry of the expression. If f has a zero at P, $\operatorname{ord}_P(1-f) = 0$ and $\langle 1-f, f\rangle = \frac{f^0}{(1-f)^{\operatorname{ord}_P f}}(P) = 1$. Finally if $\operatorname{ord}_P f < 0$, $\operatorname{ord}_P(1-f) = \operatorname{ord}_P f$ and $\langle 1-f, f\rangle = \left(\frac{f}{f-1}\right)^{\operatorname{ord}_P f}(P) = 1$.

□

Now we can state our main result which is sometimes called the "Product Formula".

Theorem 3.7.5 ("Generalized Weil reciprocity)**.** *For $f, g \in K(E)$, we have*

$$\prod_{P\in E} \langle f, g\rangle_P = 1 \ .$$

Proof. We consider g as a map from E to $\mathbb{P}_1$. If g is a constant mapping, say $g(P) = t \in K$, then $\langle f, g\rangle_P = 1/t^m$ where $m = \operatorname{ord}_P f$ so

$$\prod_{P\in E} \langle f, g\rangle_P = t^{-\Sigma \operatorname{ord}_P f} = t^0 = 1 \ .$$

Thus we can assume g is not a constant mapping. Note that $g^*(K(\mathbb{P}_1)) = K(g) \subset K(E)$. Thus if $t \in \mathbb{P}_1$, the local symbol $\langle \cdot \rangle_t$ defines a "local symbol", $\langle \cdot \rangle_\tau$ on pairs of functions in $K(g)$ by

$$\langle r \circ g, s \circ g\rangle_\tau = \langle r, s\rangle_t \, ,$$

where $r, s \in K(\mathbb{P}_1)$ are well-defined because g is onto. If $I : \mathbb{P}_1 \to \mathbb{P}_1$ is the identity map, then $g^*(I) = g$, a fact which will be useful. The theorem then follows by applying Lemma 3.7.6 to the function $r = g_*(f) = N_{K(E)/K(g)} f \circ g^{-1}$ and using Lemma 3.7.7:

$$\begin{aligned}\prod_P \langle f,g\rangle_P &= \prod_{t\in\mathbb{P}_1}\prod_{g(Q)=t}\langle f,g\rangle_Q \\ &= \prod_{t\in\mathbb{P}_1}\left\langle N_{K(E)/K(g)}f,g\right\rangle_\tau \\ &= \prod_{t\in\mathbb{P}_1}\left\langle N_{K(E)/K(g)}f\circ g^{-1},I\right\rangle_t \\ &= 1\ .\end{aligned}$$

□

Lemma 3.7.6. *For any rational function r on $\mathbb{P}_1$, we have*

$$\prod_{t\in\mathbb{P}_1}\langle r,I\rangle_t = 1\ .$$

Proof. We can write a rational function on $\mathbb{P}_1$ as

$$r = a\prod_{t\in\mathbb{P}_1}(I-t)^{n_t}$$

where $n_t = \operatorname{ord}_t r$. Since the local symbol $\langle r,I\rangle_t$ is multiplicative in r, we are reduced to the case $r=(I-t)$. Suppose $t=0$, *i.e.*, $r=I$. By ii) of the above exercise, $\langle I,I\rangle_s = 1$ for $s\neq 0,\infty$. It is easy to compute $\langle I,I\rangle_0 = -1$ and $\langle I,I\rangle_\infty = -1$, and this case follows.

Now suppose $t\neq 0$. Here we have $\langle I-t,I\rangle_s = 1$ for $s\neq 0,t,\infty$. Again it is easy to compute $\langle I-t,I\rangle_0 = -t, \langle I-t,I\rangle_t = 1/t$, and $\langle I-t,I\rangle_\infty = -1$, and we are done. □

Lemma 3.7.7. *For all $t\in\mathbb{P}_1$, we have*

$$\prod_{g(Q)=t}\langle f,g\rangle_Q = \langle N_{K(E)/K(g)}f,g\rangle_\tau \tag{3.24}$$

Proof. The proof of Lemma 3.7.7 is much more difficult and requires some preparation even to point out where the difficulty lies. We give this preparation in the next section. We can, however, show that Weil reciprocity is a consequence of Generalized Weil reciprocity. □

Corollary 3.7.8. *Let $r, s \in K(E)$ have divisors with disjoint supports, i.e., they have zeros and poles at different points of E. Then*

$$r(\operatorname{div} s) = s(\operatorname{div} r). \tag{3.25}$$

Proof. By the theorem,

$$\prod_{P \in E} \langle r, s \rangle_P = 1 \ .$$

Now $\langle r, s \rangle_P = 1$ except where r or s has a pole or zero. If r has a pole or zero at P (and hence s does not), then

$$\langle r, s \rangle_P = s(P)^{-\operatorname{ord}_P r},$$

while if s has a zero or pole at P

$$\langle r, s \rangle_P = r(P)^{\operatorname{ord}_P s} \ .$$

Since

$$r(\operatorname{div}\ s) = \prod_{P \in E} r(P)^{\operatorname{ord}_P s},$$

we see that the Generalized Reciprocity Formula says that

$$r(\operatorname{div}\ s) \cdot s(-\operatorname{div}\ r) = 1$$

which is the desired result. □

3.8 A Bit of Valuation Theory

If we examine Equation (3.24), we see that on the right we have a *global* object, while on the left side we have a product of *local* objects. We will first look at the norm $N_{K(E)/K(g)}$ and show how this can be written as a product of "local norms". To see where these "local norms" come from, we must examine the relationship between the order, ord_t, at $t \in \mathbb{P}_1$ and the orders, ord_Q at the points $Q \in E$ with $g(Q) = t$. Again it is advantageous to adopt a more abstract view if only to relate to the standard texts. By the way, references for this section are [15], Chapter 3 and [17], Chapters I and II.

The orders mentioned above are examples of discrete valuations.

Definition 3.8.1. Let K be any field. A *discrete valuation* on K is a mapping $v : K^* \to \mathbb{Z}$ such that

$$v(xy) = v(x) + v(y)$$

$$v(x+y) \geq \min(v(x), v(y)) \ .$$

We say v is *trivial* if $v(x) = 0$ for all $x \in K^*$.

The next example is very important for what follows.

Example 3.8.2. Let E be an elliptic curve over a field k. It is easy to see that ord_P is a valuation on the field $k(E)$ for any point $P \in E$. What is not so easy to see is that if k is algebraically closed, these are the only valuations on $k(E)$ that are trivial on k, the subfield of constant rational functions. This is essentially due to a famous theorem of Hilbert called the "Nullstellensatz". We indicate here how this comes about with proofs deferred to the

Appendix. Let v be a discrete valuation on $k(E)$ which is trivial on k. We want to find a point $P \in E$ such that $v = \mathrm{ord}_P$. Let $A = \{r \in k(E) : v(r) \geq 0\}$. We want to think of A as being the rational functions which are finite at our desired point P.

Exercise 3.8.3. Show that A is a *discrete valuation ring* or DVR, *i.e.*, A is a pid with a unique (nonzero) prime (or maximal) ideal $\mathfrak{m}$. Show that $\mathfrak{m} = \{r \in k(E) : v(r) > 0\}$.

The field $A/\mathfrak{m}$ is called the *residue field* of A. The form of the Nullstellensatz we use is the following:

Theorem 3.8.4 (Weak Hilbert Nullstellensatz)**.** *Let A be a ring which is finitely generated over a field k. Let $\mathfrak{m}$ be any maximal ideal of A. Then $A/\mathfrak{m}$ is an algebraic extension of k.*

See the Appendix for an indication of the proof. In our case since k is assumed to be algebraically closed, we get $A/\mathfrak{m} = k$.

There are two cases, corresponding to the cases where the desired point P is finite or infinite. First suppose $v(x) \geq 0$ where x is the usual coordinate function.

Exercise 3.8.5. Show that this implies that $v(y) \geq 0$ also.

So we have $x, y \in A$. Let $\pi : A \to A/\mathfrak{m}$ be the canonical projection, and put $a = \pi(x)$ and $b = \pi(y)$ so $a, b \in k$. It is easy to see that $P = (a, b)$ is a point on our curve E. We want to show that $v(r) = \mathrm{ord}_P\, r \quad \forall r \in k(E)$.

First note that $\pi(x) = a = x(P)$ and $\pi(y) = b = y(P)$ so $\pi(r) = r(P) \quad \forall r \in k(E)$, *i.e.*, π is the evaluation map at P. Hence

$$\mathfrak{m} = \{r \in A : r(P) = 0\} \; .$$

Now if $r \notin A, v(1/r) < 0$ so $(1/r) \in \mathfrak{m}$ and $(1/r)(P) = 0$. Hence r is not finite at P, and we see that A is the ring of rational functions which are finite at P.

Observe that v can be computed from A. Since A is a pid, we can pick a generator u of $\mathfrak{m}$.

Exercise 3.8.6. If $r \in A$ is not zero, show that we can write $r = u^d s$ with $d \in \mathbb{Z}$ and s invertible.

Now it is not hard to see that $v(r) = d$ and that u is a uniformizer at P so $\operatorname{ord}_P r = d$ too.

If $v(x) < 0$, then we take $P = \mathcal{O}$. We leave the details of this case to the "interested reader".

Closely related to discrete valuations are absolute values. In fact McCarthy's valuations are Serre's absolute values.

Definition 3.8.7. Let K be any field. An *absolute value* on K is a real-valued function $x \mapsto |x|$ on K which satisfies the following conditions:

(i) $|x| \geq 0$ and $|x| = 0$ if and only if $x = 0$,

(ii) $|xy| = |x| \cdot |y|$,

(iii) $|x + y| \leq \max(|x|, |y|)$.

Exercise 3.8.8.

(i) Let v be a discrete valuation on K and $a \in \mathbb{R}$ be between 0 and 1. Set $|x| = a^{v(x)}$. Show that $| \cdot |$ is an absolute value on K.

(ii) Show $|-x| = |x|$.

(iii) Suppose $|x| > |y|$. Show $|x + y| = |x|$.

It is *not* true that every absolute value on K comes from a discrete valuation, however we make the following additional assumptions on our absolute values which will insure that they come from a discrete valuation. We assume that $|K^*|$, the set of absolute values of all elements of K^*, is a discrete subset of $\mathbb{R}^*$.

Exercise 3.8.9. Let $|\cdot|$ be an absolute value on K. Show that there is $\lambda > 0$ in $\mathbb{R}$ such that $v(x) = \lambda \cdot \exp(|x|)$ is a discrete valuation on K.

Remark. The terminology regarding valuations, absolute values and the like is a mess. Various books use different names for the same concepts, and the same names for different concepts. There are actually *four* concepts which are more or less equivalent, valuations, absolute values, valuation rings, and places. We have already seen valuations and absolute values. If R is a subring of a field K we say R is a *valuation ring* for K if $x \in K$ and $x \notin R$ implies $x^{-1} \in R$. A *place* of K is a "homomorphism" from K into the set consisting of the elements of some field F and another element called ∞''. By this we mean that the intuitive algebraic rules regarding ∞ are followed, *e.g.*, $x \pm \infty = \infty, 1/0 = \infty, 0 \cdot \infty$ is undefined, *etc.*

If we have a valuation v, we get a valuation ring by setting

$$R = \{x \in K : v(x) \geq 0\} \ .$$

Let $\mathfrak{m} = \{x \in R : v(x) = 0\}$. We get a place by mapping $x \in R$ into $R/\mathfrak{m}$ and if $x \in K$ is not in R we map it to ∞.

The notion of valuation is confused because some people consider valuations whose values lies in more general groups than others. There are various characterizations of all of these concepts that correspond to *discrete* valuations, *e.g.*, in a valuation ring corresponding to a discrete valuation the ideal $\mathfrak{m}$ is *principal.*

In addition, there are different versions of rule iii) in the definition of absolute value. If one merely requires the triangle inequality

$$|x + y| \leq |x| + |y|$$

we get a somewhat more general notion. The ones we consider are called *ultrametric* by the French and *non-archimedean* by everybody else. Roughly speaking, the only other ones (the archimedean ones) are the usual absolute values on the reals and the complexes. [10] is a good place to try and get this all straightened out.

We have tried to avoid valuation rings and places, and have used mostly valuations in the body of these notes and absolute values in the Appendix.

Let $|\cdot|$ be an absolute value on K. Recall the construction of the real numbers from the rational numbers. One can copy that construction on K to get the *completion* $\hat{K}$. One defines cauchy sequences and limits just as in the rational case. Thus $\hat{K}$ is an extension field of K in which every cauchy sequence converges. The following exercise is not difficult although it is a little fussy:

Exercise 3.8.10. If $\sigma : K \to L$ is an isomorphism of fields with absolute values, then we say σ is *analytic* if $|x|_K = |\sigma(x)|_L$. Suppose L is an extension of K which happens to be complete. Show there is an analytic isomorphism from $\hat{K}$ to a subfield of L which is the identity on K, *i.e.*, $\hat{K}$ is analytically K-isomorphic to a subfield of L. In particular, this shows that if K' is a complete extension of K such that every element of K' is the limit of some cauchy sequence of elements of K, then K' is analytically K-isomorphic to $\hat{K}$.

Exercise 3.8.11. Let E be an elliptic curve and $P \in E$. Let

$$A = \{r \in k(E) : r(P) < \infty\} .$$

Let $\hat{A}$ be the completion of A with respect to the valuation $\mathrm{ord}_{P'}$. Show that $\hat{A}$ is isomorphic to $k[[u]]$, the ring of formal power series in a uniformizer u at P. (This situation is investigated in the case $P = \mathcal{O}$ in [6], Chapter IV.)

If L is an extension of K and K has a valuation (and an associated absolute value), then there may be many ways to extend the valuation on K to L. In the finite separable case they are classified by the following theorem:

Theorem 3.8.12. *Let $L = K(x)$ be a finite separable extension of K, and let v be a valuation on K. Let m_x be the minimal polynomial of x. Suppose m_x factors into r nonconstant irreducible factors when considered as a polynomial in $\hat{K}[X]$. Then v has exactly r distinct extensions to L.*

Furthermore, if we let V be an extension of v to L which corresponds to the factor f of m_x in $\hat{K}[X]$, and let $\hat{L}$ be the completion of L with respect to V, then there is $\hat{x} \in \hat{L}$ with $\hat{L} = \hat{K}(\hat{x})$ and f is the minimal polynomial of $\hat{x}$ over $\hat{K}$.

The purely inseparable case is handled by the following:

Proposition 3.8.13. *Let L be a purely inseparable extension of K. Then any valuation on K has a unique extension to L.*

We defer the proof of Theorem 3.8.12 to the Appendix. Note that the hypothesis that L be generated by a single element over K is not a restriction since every finite separable extension is generated by a single element (see any algebra book, say [2], page 185). For the proof of the proposition we note that if L is a purely inseparable extension of K, then there is an integer $e > 0$ with the property that $x^{p^e} \in K$ for all $x \in L$. This enables us to write a formula for the extension of a valuation on K to L. We leave the details as an exercise. As in Section 3.5, we can put the theorem and the proposition together to cover the case of an arbitrary finite extension. We can now use these results to prove the desired result concerning the norm.

Theorem 3.8.14. *Let L be a finite extension of K. Let v be a valuation on K, and let $v_1, v_2, \ldots, v_r$ be the various extensions of v to L. Let $\hat{K}$ be the completion of K and $\hat{L}_i$ be the completion of L with respect to v_i. Then for all $y \in L$ we have*

$$N_{L/K}(y) = \prod_{i=1}^{r} N_{\hat{L}_i/\hat{K}}(y) \ .$$

Proof. First we assume that L is a separable extension so we can write $L = K(x)$ for some $x \in L$. By the results of Section 3.5, $N_{L/K}(x)$ is plus or minus the constant term of m_x. This constant term is the product of the constant terms of its factors in $\hat{K}[X]$ which, in turn, are plus or minus the various $N_{\hat{L}_i/\hat{K}}(x)$. Hence the theorem holds for the particular element x.

Fix $y \in L^*$. (If $y = 0$, the result is trivial.) Consider the fields $K(y \cdot (x + z))$ where z runs through K. Since v is a nontrivial valuation, K must have infinitely many elements. Since L is separable over K, there are only finitely many fields between L and K ([2], page 185 again). Hence for some distinct $z, z' \in K$ we have $K(y \cdot (x + z)) = K(y \cdot (x + z'))$. Call this intermediate field K'.

We must have $y. (z \ - z') \in K'$. Since $z - z' \in K^*$, we must have $y \in K'$. Thus $y \cdot (x+z) \in K'$ so $yx \in K'$, and since $y \neq 0$, we must have $x \in K'$. Therefore $L = K(y \cdot (x+z))$. Since $z \in K$, we also have $L = K(x+z)$. Hence we have the theorem for the particular elements $y \cdot (x+z)$ and $x+z$ of L. Finally we get

$$\begin{aligned} N_{L/K}(y) &= \frac{N_{L/K}(y \cdot (x+z))}{N_{L/K}(x+z)} \\ &= \frac{\prod_{i=1}^{r} N_{\hat{L}_i/\hat{K}}(y \cdot (x+z))}{\prod_{i=1}^{r} N_{\hat{L}_i/\hat{K}}(x+z)} \\ &= \prod_{i=1}^{r} \frac{N_{\hat{L}_i/\hat{K}}(y \cdot (x+z))}{N_{\hat{L}_i/\hat{K}}(x+z)} \\ &= \prod_{i=1}^{r} N_{\hat{L}_i/\hat{K}}(y) \ . \end{aligned}$$

This finishes the separable case.

As usual we can now consider a field M with M separable over K and L purely inseparable over M. In the purely inseparable case Equation (3.18) shows that $N(x) = x^{p^e}$ where $p^e = [L : M]$. It is not hard to show that the degree does not change under completion with respect to any valuation on M (and its unique extension to L). We again leave the details as an exercise. □

3.9 Completion of the Proof of Weil Reciprocity

We now complete the proof of generalized Weil reciprocity by proving Lemma 3.7.7. We had an elliptic curve E and rational functions

$f, g \in K(E)$. We considered g as a rational mapping to the projective line $\mathbb{P}_1$. Then for $t \in \mathbb{P}_1$ we wanted to prove the following formula:

$$\prod_{g(Q)=t} \langle f, g \rangle_Q = \left\langle N_{K(E)/K(g)} f, g \right\rangle_\tau$$

where $\langle \cdot \rangle_\tau$ is defined by

$$\langle r \circ g, s \circ g \rangle_\tau = \langle r, s \rangle_t \ .$$

Let $K_g = \widehat{K(g)}$ be the completion of this field with respect to the valuation $\operatorname{ord}_\tau(h \circ g) = \operatorname{ord}_t h$ for $h \in K(\mathbb{P}_1)$. Then ord_τ extends uniquely to K_g.

Now if $v, w \in K_g$, the local symbol $\langle v, w \rangle_\tau$ can be defined as the limit of the local symbols in $K(g)$. It is easy to see that the values of the local symbols will converge because the difference of two functions that are close in this metric must have a zero at t. This "completed" local symbol is still multiplicative in v and w. Similarly for $Q \in E$, we can extend the local symbol to $K_Q = \widehat{K(E)_Q}$, the completion of $K(E)$ with respect to the valuation ord_Q.

Proposition 3.1.4 tells us that $\operatorname{ord}_Q = e \cdot \operatorname{ord}_\tau$ on $K(g)$ where $e = e_g(Q)$ is the ramification index of g at $Q \in E$. Consider the extensions of the valuation $e \cdot \operatorname{ord}_\tau$ on K_g to K_Q. The following theorem describes the situation completely.

Theorem 3.9.1. *Let L be a field which is complete with respect to some valuation v, and let M be a finite extension of L. Then there is a unique extension of v to M. In fact, if $x \in M$, then*

$$v(x) = (1/d)v(N_{M/L}x)$$

where $d = [M : L]$.

This result is a translation of Theorem 4.3.1 in the Appendix. We know that ord_Q is an extension of $e \cdot \mathrm{ord}_\tau$. Hence if we write $d = [K_Q : K_g]$ and $N_Q = N_{K_Q/K_g}$, then

$$\mathrm{ord}_Q(r) = (e/d)\,\mathrm{ord}_\tau(N_Q r) \text{ for } r \in K_Q \ .$$

We now show $e = d$ where $e = e_g(Q)$ and $d = [K_Q : K_g]$. In fact, we will show that $1,\ f, f^2, \dots,\ f^{e-1}$ forms a basis for K_Q as a vector space over K_g. First assume $t = g(Q) = 0$ so $e = \mathrm{ord}_Q\, g$. Let $r \in K_Q$ and $a_0 = r(Q)$. Suppose

$$\mathrm{ord}_Q(r - a_0) = h_1 = k_1 e + p_1 \text{ with } 0 \le P_1 < e \ .$$

Then we can write $r - a_0 = s_1 f^{\ell_1} g^{k_1}$ where $s_1(Q) = a_1 \neq 0$. We now set $r_1 = a_0 + a_1 f^{\ell_1} g^{k_1}$, and $\mathrm{ord}_Q(r - r_1) = h_2 > h_1$. We can set $h_2 = k_2 e + \ell_2$ and continue this process producing a sequence of functions $\{r_i\}$ which converges to r and is of the form

$$r_i = q_0(g) + q_1(g) f + \cdots + q_{e-1}(g) f^{e-1}$$

where each q_j is a polynomial. This proves our result ($e = d$) in this case. If $g(Q) \neq 0, \infty$, substitute $g - g(Q)$ for g in the above construction. If Q is a pole of g, use $1/g$ for g. Hence we have shown that

$$\mathrm{ord}_Q(r) = \mathrm{ord}_\tau(N_Q r) \text{ for } r \in K_Q. \tag{3.26}$$

Theorem 3.8.14 tells us that

$$N_{K(E)/K(g)} f = \prod_{g(Q)=t} N_Q f \ .$$

Thus the right-hand side on our equation satisfies

$$\left\langle N_{K(E)/K(g)} f, g \right\rangle_\tau = \prod_{g(Q)=t} \left\langle N_Q f, g \right\rangle_\tau ,$$

and it suffices to prove the following:

Claim. $\langle f, g\rangle_Q = \langle N_Q f, g\rangle_\tau$.

Since everything in sight is multiplicative, we can assume that f is a uniformizer at Q *i.e.*, $\operatorname{ord}_Q f = 1$. On the other hand if $\operatorname{ord}_Q g = 0$, then

$$\langle f, g\rangle_Q = g(Q)^{-\operatorname{ord}_Q f} = 1/t,$$

while

$$\langle N_Q f, g\rangle_\tau = t^{-\operatorname{ord}_\tau N_Q f} = 1/t$$

since $\operatorname{ord}_\tau N_Q f = \operatorname{ord}_Q f = 1$ by (3.26). Hence we can assume $\operatorname{ord}_Q g \neq 0$ or that $t = 0$ or ∞.

Since $g = g^*(I)$, we must have $\operatorname{ord}_\tau g = \pm 1$. Suppose $\operatorname{ord}_\tau g = 1$. By the definition of the local symbol

$$\begin{aligned}\langle N_Q f, g\rangle_\tau &= -\frac{(N_Q f)\circ g^{-1}}{I}(t)\\ &= \frac{N_Q f}{g}(Q)\end{aligned}$$

where the expression $N_Q f(Q)$ must be interpreted with a grain of salt since $N_Q f$ lies in the completion of $K(g)$ and thus may not be a proper function. As we have remarked above, however, there are functions in $K(g)$ near $N_Q f$, and if they are near enough, they all have the same value which we take for $N_Q f(Q)$.

On the other hand, $\operatorname{ord}_Q g = e \cdot \operatorname{ord}_t I = e$ so

$$\langle f, g\rangle_Q = (-1)^e \frac{f^e}{g}(Q) \tag{3.27}$$

Therefore in this case it suffices to show that the function $N_Q f / f^e$ takes the value $(-1)^{e-1}$ at the point Q. If we have $t = \infty$, then a similar argument shows that it suffices to prove the exact same thing.

Let the minimal polynomial of f over K_g be given by

$$m_f(X) = X^e + a_1 X^{e-1} + \cdots + a_e$$

where the a_i are in K_g. By the definition of the norm and the proof of Theorem 3.5.7, we get that $a_e = (-1)^e N_Q f$. Now consider the equation $m_f(f) = 0$, *i.e.*,

$$f^e + a_1 f^{e-1} + \cdots + a_e = 0. \tag{3.28}$$

If $a_i \neq 0$, then

$$\begin{aligned} \operatorname{ord}_Q(a_i f^{e-i}) &= \operatorname{ord}_Q a_i + \operatorname{ord}_Q f^{e-i} \\ &= e \cdot \operatorname{ord}_\tau a_i + e - i \\ &\equiv -i \ (\bmod\ e)\ . \end{aligned}$$

Hence the nontrivial monomials of Equation (3.28) all have distinct orders at Q except possibly for f^e and a_e. Thus the only way for all of the monomials to sum to 0 is for $\operatorname{ord}_Q f^e$ to equal $\operatorname{ord}_Q a_e$. Also since

$$\operatorname{ord}_Q(f^e \ + a_e) \geq \operatorname{ord}_Q f^e = \operatorname{ord}_Q a_e \ ,$$

all of the other monomials must have higher order at Q, *i.e.*,

$$\operatorname{ord}_Q f^e = \operatorname{ord}_Q a_e < \operatorname{ord}_Q(a_i f^{e-i}) \text{ for } 1 \leq i \leq e-1\ .$$

Hence since

$$\operatorname{ord}_Q\ (f^e \ + a_e) = \operatorname{ord}_Q(-a_1 f^{e-1} - a_2 f^{e-2} - \cdots - a_{e-1} f)$$

$$\geq \min_{1\leq i\leq e-1}\{a_i f^{e-i}\}$$

$$> \operatorname{ord}_Q f^e = \operatorname{ord}_Q a_e,$$

after dividing by f^e, we get

$$\operatorname{ord}_Q(1 + a_e/f^e) > \operatorname{ord}_Q 1 = 0\ .$$

This tells us that

$$\frac{a_e}{f^e}(Q) = -1$$

so

$$\frac{N_Q f}{f^e} Q = (-1)^{e-1},$$

and we have finished the proof of Lemma 3.7.7 and of generalized Weil reciprocity.

3.10 The Weil Pairing

We end these notes with an application of generalized Weil reciprocity to the Weil pairing. Definition 2.4.5 of the Weil pairing is essentially the one given in [6], page 96. In Exercise 3.16 on page 108, Silverman gives an alternative definition which involves functions of much lower degree. In [13] a similar definition is used, but it is slightly incorrect. In this section we give the correct version of the definition in [13] and prove it is equivalent to Definition 2.4.5.

Let E be an elliptic curve over an algebraically closed field K. Pick $N > 0$ such that N is prime to the characteristic of K. Let $P, Q \in E[N]$, the subgroup of E of N-torsion points. Suppose $P, Q \neq \mathcal{O}$ and $P \neq Q$. Fix $P', Q' \in E[N^2]$ such that

$$P = N \cdot P' \text{ and } Q = N \cdot Q' \ .$$

Pick functions f_P and $g_P \in K(E)$ such that

$$\begin{aligned} \operatorname{div} f_P &= N\langle P\rangle - N\langle \mathcal{O}\rangle & (3.29) \\ \operatorname{div} g_P &= [N]^*(\langle P\rangle - \langle \mathcal{O}\rangle) \\ &= \sum_{Q\in E[N]} \langle P' + Q\rangle - \langle Q\rangle \\ && (3.30) \end{aligned}$$

where $[N]$ is the rational mapping multiplication by N. Then

$$\operatorname{div}(f_P \circ [N]) = [N]^*(\operatorname{div} f_P) = N([N]^*(\langle P\rangle - \langle \mathcal{O}\rangle)) = N \operatorname{div}(g_P)$$

therefore we may additionally assume that

$$\begin{aligned} g_P^N &= f_P \circ [N], \quad \text{and} & (3.31) \\ g_Q^N &= f_Q \circ [N] \end{aligned}$$

Pick f_Q and g_Q similarly. In definition 2.4.5 (and in [6]), the Weil pairing was defined by

$$w(P, Q) = \frac{g_Q \circ \tau_P}{g_Q}$$

where τ_S is translation by S. Recall that the function on the right-hand side of the above equation is multiplication by an N^{th}-root of unity, and by an "abuse of terminology" we define $w(P, Q)$ to be this root.

The next theorem gives the correct version of the definition in [13].

Theorem 3.10.1. *Let $P, Q \in E[N]$. Then*

$$w(P,Q) = (-1)^N \frac{f_P(Q)}{f_Q(P)} \cdot \left[\frac{f_Q}{f_P}\right](\mathcal{O}) \ .$$

Proof. Let h be any rational function with

$$\operatorname{div} h = (N-1)\langle Q'\rangle + \langle Q' - Q\rangle - N\langle \mathcal{O}\rangle \ .$$

By Generalized Weil reciprocity, we have

$$\prod_{S \in E} \langle g_P, h\rangle_S = 1 \ .$$

The only nontrivial contributions in the above product come from $S = Q', Q' - Q$ and the zeros and poles of g_P.

First consider $S = Q'$ and $Q' - Q$,

$$\begin{aligned} \langle g_P, h\rangle_{Q'} &= g_P^{N-1}(Q') \text{ and} \\ \langle g_P, h\rangle_{Q'-Q} &= g_P(Q'-Q) \text{ so} \\ \langle g_P, h\rangle_{Q'} \cdot \langle g_P, h\rangle_{Q'-Q} &= \frac{g_P(Q'-Q)}{g_P(Q')} \cdot g_P^N(Q') \\ &= \frac{g_P(S)}{g_P(S+Q)} \cdot f_P \circ [N](Q') \\ &= \frac{f_P(Q)}{w(P,Q)} \end{aligned}$$

where we have put $S = Q' - Q$ and used (3.31).

Now let

$$H(S) = \prod_{T \in E[N]} h(S+T)$$

so

$$\begin{aligned}
\operatorname{div} H &= \sum_{T\in E[N]} \operatorname{div}(h\circ\tau_T) = \sum_{T\in E[N]} \tau_{T^*}(\operatorname{div}\ h) \\
&= \sum_{T\in E[N]} (N-1)\left\langle Q'-T\right\rangle + \left\langle Q'-Q-T\right\rangle - N\left\langle -T\right\rangle \\
&= \sum_{T\in E[N]} N\left\langle Q'-T\right\rangle - N\left\langle -T\right\rangle \\
&= [N]^*(N\left\langle Q\right\rangle - N\left\langle \mathcal{O}\right\rangle) \\
&= \operatorname{div}(f_Q\circ[N])\ .
\end{aligned}$$

where we used the fact that $T \mapsto Q+T$ is a permutation of $E[N]$ since $Q \in E[N]$. Since we had the usual freedom in the choice of h, we can assume

$$H = f_Q \circ [N] = g_Q^N$$

by (3.31).

We now consider the zeros of g_P, *i.e.* $\{P'+R \mid R\in E[N]\}$.

$$\begin{aligned}
\prod_{\operatorname{ord}_S(g_P)>0} \left\langle g_P, h\right\rangle_S &= \prod_{T\in E[N]} \frac{1}{h(P'+T)} \\
&= \frac{1}{H(P')} \\
&= \left\langle g_P, H\right\rangle_{P'}\ .
\end{aligned}$$

For the poles of g_P, we have

$$\begin{aligned}\prod_{\mathrm{ord}_S(g_P)<0} \langle g_P, h\rangle_S &= \langle g_P, h\rangle_{\mathcal{O}} \cdot \prod_{T\in E[N]-\mathcal{O}} h(T) \\ &= (-1)^{1\cdot N} \frac{g_P^{-N}}{h^{-1}}(\mathcal{O}) \cdot \prod_{T\in E[N]-\mathcal{O}} h(T) \\ &= (-1)^{1\cdot N} \frac{g_P^{-N}}{H^{-1}}(\mathcal{O}) \\ &= \langle g_P, H\rangle_{\mathcal{O}} \ .\end{aligned}$$

Putting this all together, we get

$$\begin{aligned}1 &= [\langle g_P, h\rangle'_Q \cdot \langle g_P, h\rangle_{Q'-Q}] \cdot \langle g_P, H\rangle_{P'} \cdot \langle g_P, H\rangle_{\mathcal{O}} \\ &= \frac{f_P(Q)}{w(P,Q)} \cdot \frac{1}{H(P')} \cdot (-1)^N \frac{H}{g_P^N}(\mathcal{O}) \\ &= \frac{f_P(Q)}{w(P,Q)} \cdot \frac{1}{f_Q(N\cdot P')} \cdot (-1)^N \frac{f_Q\circ[N]}{f_P\circ[N]}(\mathcal{O}) \\ &= \frac{1}{w(P,Q)} \cdot \frac{f_P(Q)}{f_Q(P)} \cdot (-1)^N \frac{f_Q}{f_P}(\mathcal{O}) \ .\end{aligned}$$

□

If you examine the choices made for f_P, f_Q, g_P, and g_Q, you will see that f_P and f_Q are really completely arbitrary (with the given divisor) so we can pick them so that

$$\frac{f_Q}{f_P}(\mathcal{O}) = 1 \ .$$

For this choice the Weil pairing can be written

$$w(P,Q) = (-1)^N \frac{f_P(Q)}{f_Q(P)} \ .$$

This has appeared a number of times in the literature incorrectly, without the $(-1)^N$, *e.g.* [13].

Chapter 4

Appendix

4.1 The Nullstellensatz

In Section 3.8 we used a famous theorem due to Hilbert called the "Nullstellensatz". This theorem has a number of forms, and the one we used is usually called the "weak" Nullstellensatz. The Nullstellensatz is a basic theorem for the study of algebraic geometry, and most books on elementary algebraic geometry contain a proof. One way to state it is that a maximal ideal of polynomials in many variables with coefficients in an algebraically closed field have a common zero. A straightforward proof of this version can be found in [14]. Another elementary proof is in [11]. We follow the treatment in [12].

In our situation we have a field k and a ring R which is finitely generated over k, *i.e.*, $R = k[x_1, x_2, \ldots, x_n]$. Now to make the theorem interesting, the x_i must be transcendental over k, but they may not be algebraically independent over k, *i.e.*, R may not be the polynomial ring in n variables over k. The theorem tells us that if we divide R by a maximal ideal $\mathfrak{m}$, the residue field $R/\mathfrak{m}$ *is* algebraic over k. So roughly speaking, $\mathfrak{m}$ must contain all of the transcendence.

The proof proceeds by first showing that $A = R/\mathfrak{m}$ contains a subring B which is a polynomial ring over k in perhaps fewer than n variables, and that A is integral over B, *i.e.*, every element of A satisfies a monic polynomial with coefficients in B. This result is called Noether's Normalization Theorem. Before we get to its proof, we need some elementary results concerning integral elements.

Proposition 4.1.1. *Let B be a subring of a ring A and let $\alpha \in A$. The following properties are equivalent:*

(i) α is integral over B.

(ii) $B[\alpha]$ is finitely generated as a B-module where $B[\alpha]$ is the subring of A generated by B and α.

(iii) There exists a subring B_1 with $B \subset B_1 \subset A$ and $\alpha \in B_1$ such that B_1 is finitely generated as a B-module.

Proof. i) $\Rightarrow$ (ii) Since α is integral over B, there exists $F(X) \in B[X]$ such that

$$F(X) = X^n + b_1 X^{n-1} + \cdots + b_n \ ,$$

and $F(\alpha) = 0$. Hence

$$\alpha^n = -\sum_{i=1}^{n} b_i \alpha^{n-i}$$

so $B[\alpha]$ is generated by 1, $\alpha, \alpha^2, \ldots, \alpha^{n-1}$ as a B-module.

ii) $\Rightarrow$ (iii) is trivial.

iii) $\Rightarrow$ (i). Let $b_1, b_2, \ldots, b_n \in B_1$ be a set of generators for B_1 as a B-module. Since $\alpha \in B_1, \alpha b_i \in B_1$ for $i = 1, 2, \ldots, n$ so there are elements $\beta_{i,j} \in B$ such that

$$\alpha b_i = \sum_{j=1}^{n} \beta_{i,j} b_j \text{ for } i = 1, 2, \ldots, n. \tag{4.1}$$

Let F be the polynomial in $B_1[X]$ defined by

$$F(X) = \det(X \cdot I_n - [\beta_{i,j}])$$

where I_n is the $n \times n$ identity matrix, and $[\beta_{i,j}]$ is the obvious thing. Equation (4.1) implies that $F(\alpha)b_i = 0$ for $i = 1, 2, \ldots, n$. Hence $F(\alpha) = 0$. Since the coefficients of F are in B and F is clearly monic, this shows α is integral over B. □

Now we give a few corollaries.

Corollary 4.1.2. *If $\alpha_1, \alpha_2, \ldots, \alpha_n \in A$ are integral over B, then $B[\alpha_1, \alpha_2, \ldots, \alpha_n]$ is finitely generated as a B-module and is integral over B, i.e., every element of $B[\alpha_1, \ldots, \alpha_n]$ is integral over B.*

Proof. The proof is by induction on n. For $n = 1$, this is i) $\Rightarrow$ ii) of the proposition. Assume the desired result holds for $n - 1$. Then $B_1 = B[\alpha_1, \ldots, \alpha_{n-1}]$ is finitely generated over B so

$$B_1 = \sum_{i=1}^{s} \beta_i B$$

for some s and some $\beta_i \in B_1$. Since α_n is integral over B,

$$B[\alpha_n] = \sum_{j=0}^{m-1} \alpha_n^j B$$

for some m. Hence

$$\begin{aligned} B[\alpha_1, \ldots, \alpha_n] &= B_1[\alpha_n] \\ &= \sum_{j=0}^{m-1} \alpha_n^j B_1 \\ &= \sum_{i,j} \alpha_n^j \beta_i B \end{aligned}$$

which is a finitely generated B-module. □

Corollary 4.1.3. *Let $B_A = \{\alpha \in A : \alpha \text{ is integral over } B\}$. Then B_A is a subring of A.*

Proof. Take $\alpha, \beta \in B_A$. By Corollary 4.1.2, $B[\alpha, \beta]$ is integral over B so $\alpha \pm \beta$ and $\alpha\beta$ which are members of $B[\alpha, \beta]$ are integral over B. □

Corollary 4.1.4. *If A is integral over B and B is integral over C, then A is integral over C.*

We leave this proof as an exercise since it is an easy consequence of Corollary 4.1.4.

Corollary 4.1.5. *If $\alpha \in A$ and α is integral over B_A, then $\alpha \in B_A$.*

This proof is an easy consequence of the previous corollary.

Now we can prove the Normalization Theorem.

Theorem 4.1.6 (Noether's Normalization Theorem)**.** *Let A be an integral domain which is finitely generated over a field, i.e., $A = k[x_1\ x_2, \ldots, x_n]$. Then there are elements $y_1, y_2, \ldots, y_r \in A$ such that*

(i) the subring $B = k[y_1, \ldots, y_r]$ is isomorphic (as an algebra over k) to the ring of polynomials in r variables, i.e., $y_1, y_2, \ldots, y_r$ are algebraically independent over k,

and

(ii) A is integral over B.

Proof. The proof is by induction on n. If $n = 0$ there is nothing to prove. Let $X_1, X_2, \ldots, X_n$ be indeterminates, and $R = k[X_1, \ldots, X_n]$. Define a surjective k-homomorphism $\varphi : R \to A$ by $\varphi(X_i) = x_i$ for $i = 1, 2, \ldots, n$. Since A is an integral domain, $\mathfrak{p} = \ker \varphi$ is a prime ideal of R. If $\mathfrak{p} = 0$, $A = R$, and we are done. Let $F \in \mathfrak{p} - 0$. If $F \in K$, then $F \cdot F^{-1} = 1 \in \mathfrak{p}$ so $\mathfrak{p} = R$, and $A = 0$, and we are done. Thus we can assume F is nonconstant. We use lemmas 4.1.7 and 4.1.8 below.

By Lemma 4.1.7, R is integral over S so $R/\mathfrak{p}$ is integral over $A_1 = S/S \cap \mathfrak{p}$ by Lemma 4.1.8. Since $F \in S \cap \mathfrak{p}$, the subring $A_1 \subset R/\mathfrak{p}$ is generated mod $(S \cap \mathfrak{p})$ over k by the G_i for $i = 2, 3, \ldots, n$. By the induction hypothesis applied to A_1 there are $y_1, y_2, \ldots, y_r \in A_1$ such that $y_1, y_2, \ldots, y_r$ are algebraically independent over k, and A_1 is integral over $B = k[y_1, \ldots, y_r]$. By Corollary 4.1.4 since $A = R/\mathfrak{p}$ is integral over A_1 and A_1 is integral over B, A is integral over B. □

Lemma 4.1.7 (Nagata)**.** *Let R be a polynomial ring, $k[X_1, \ldots, X_n]$, over a field k, and let F be a nonconstant polynomial in R. Then* *there exist integers $m_2, m_3, \ldots, m_n \geq 0$ such that R is integral over the subring $S = k[F, G_2, G_3, \ldots, G_n]$ where G_i is the polynomial*

$$G_i = X_i - X_1^{m_i} \ ,$$

Proof. Clearly if we adjoin X_1 to S, we can get the rest of the X_i's so $R = S[X_1]$. If we can show that X_1 is integral over S for an appropriate choice of $m_2, m_3, \ldots, m_n$, then by Proposition 4.1.1, R will be integral over S, and we will be done.

Let $m_2, m_3, \ldots, m_n > 0$ and put

$$G_i = X_i - X_1^{m_i} \text{ for } i = 2, 3, \ldots, n \ .$$

Let T be an indeterminate over S so $S[T]$ is the polynomial ring in one variable over S. Recall that $F \in R$ so F is a polynomial in n variables. Define $H(T) \in S[T]$ by

$$H(T) = F(T, G_2 + T^{m_2}, \ldots, G_n + T^{m_n}) - F(X_1, \ldots, X_n) \ .$$

Then $H(X_1) = 0$. Hence to show that X_1 is integral over S, it suffices to show that we can pick $m_2, m_3, \ldots, m_n$ such that the leading coefficient of H is independent of $G_2, G_3, \ldots, G_n$ since then this leading coefficient will have to be a constant *i.e.* in k, and we can divide it out.

Let $m_1 = 1$, and let δ be the total degree of F. Now F is a linear combination of monomials

$$M_\alpha(X_1, \ldots, X_n) = X_1^{\alpha_1} X_2^{\alpha_2} \cdots X_n^{\alpha_n}$$

where $\alpha = (\alpha_1, \ldots, \alpha_n)$ with $\alpha_i \geq 0$. That is

$$F(X_1, \ldots, X_n) = \sum_{|\alpha| < \delta} a_\alpha M_\alpha(X_1, \ldots, X_n)$$

with $|\alpha| = \Sigma_{i=1}^{n} \alpha_i$ and $a_\alpha \in k$. Hence

$$\begin{aligned} H(T) = \sum_{|\alpha|<\delta} & a_\alpha M_\alpha(T^{m_1}, G_2 + T^{m_2}, \ldots, G_n + T^{m_n}) - \\ & F(X_1, \ldots, X_n) \\ = \sum_{|\alpha|<\delta} & a_\alpha \cdot (T^{\Sigma_{i=1}^{n} \alpha_i m_i} + H_\alpha(T)) - F(X_1, \ldots, X_n) \end{aligned}$$

where the $H_\alpha(T) \in S[T]$ are such that

$$\deg_T H_\alpha(T) < \sum_{i=1}^{n} \alpha_i m_i,$$

i.e., we have collected the factors in M_α which contain just powers of T and the remaining terms (the H_α) have lower degree in T. Hence the only way that the leading coefficient of H can involve any X_i's is that some of the $T^{\Sigma_{i=1}^{n} \alpha_i m_i}$ cancel for different values of α. So if we pick $m_2, m_3, \ldots, m_n$ so the $\Sigma_{i=1}^{n} \alpha_i m_i$ are all different for all α with $|\alpha| < \delta$, we will be done. There are many possibilities; $m_i = (\delta + 1)^{i-1}$ works, for example. □

Lemma 4.1.8. *Let U be a subring of some ring V and $\mathfrak{a}$ an ideal of V. Then $U/\mathfrak{a} \cap U$ can be identified with a subring of $V/\mathfrak{a}$. If V is integral over U, then $V/\mathfrak{a}$ is integral over $U/\mathfrak{a} \cap U$.*

Proof. Easy, left as exercise. □

Now we can easily prove the main result of this section.

Theorem 4.1.9 (Weak Hilbert Nullstellensatz)**.** *Let R be a ring which is finitely generated over a field k. Let $\mathfrak{m}$ be any maximal ideal of R. Then $R/\mathfrak{m}$ is an algebraic extension of k.*

Proof. Since R is finitely generated over k, so is $A = R/\mathfrak{m}$, and since A is a field, it certainly is an integral domain so the Normalization Theorem 4.1.6 applies. Thus there are elements $y_1, y_2, \ldots, y_r \in A$, algebraically independent over k such that A is integral over $B = k[y_1, \ldots, y_r]$. If $r > 0$, then since A is really a field $\eta = y_1^{-1} \in A$. Hence η is integral over a polynomial ring, namely B. Therefore η satisfies

$$\eta^m + F_1\eta^{m-1} + \cdots + F_m = 0$$

for some $F_i \in B = k[y_1, \ldots, y_r]$. If we multiply by y_1^m, we get

$$1 + F_1 y_1 + \cdots + F_m y_1^m = 0$$

which contradicts the fact that the $y_1, y_2, \ldots, y_m$ are algebraically independent over B. Therefore $r = 0$, and A is algebraic over k. □

4.2 Newton's Polygon

In Section 8 we needed to know the different ways a valuation v on a field K could be extended to a finite separable extension $L = K(x)$. The answer turned out to depend on the factorization of the minimal polynomial, m_x, over the completion of K. In the next sections we indicate the proof of this result (Theorem 3.8.12).

As you may expect, the main difficulty is to find *one* extension of v to L. There are a number of approaches to this problem. One is to assume that K is complete. Then it is not hard to show that if an extension exists, it must satisfy

$$v(x) = (1/e)v(N_{L/K}x) \quad \forall x \in L \tag{4.2}$$

where e is the degree of L over K. One can then use this formula to *define* the extension, and it remains to show the extension is, indeed, a valuation on L. This can be done by means of Hensel's Lemma (see [8] or [18], for example) or by using "Newton's Polygon", (see [10]). The incomplete case then follows easily. Another way is to look at the associated valuation ring of v, and use Zorn's Lemma to find a valuation ring of L which extends it (see [9] or [15]).

All of these methods have the advantage (or disadvantage) of working for more general valuations than discrete ones. The treatment in [17] applies only to discrete valuations, but avoids the use of difficult lemmas. This treatment makes heavy use of the notion of integrality and develops some of the theory of Dedekind Domains to get at the extension results. Although in some abstract sense, Serre's approach may be the "best" for our purpose, we have decided to use Cassels's treatment because it is more elementary and computational in spirit.

Let K be a field with a valuation v. Let $|\cdot|$ be the associated absolute value, *i.e.*, $|x| = a^{|v(x)|}$ for some $a \in [0,1]$. We are first interested in extending the absolute value to the field $K(X)$. Fix $C > 0$. For

$$f(X) = a_0 + a_1 X + a_2 X^2 + \cdots + a_n X^n \in K[X]$$

define

$$\|f\| = \|f\|_C = \max_j C^j |a_j| \ .$$

For $h = f/g \in K(X)$, put $\|h\| = \|f\|/\|g\|$.

Exercise 4.2.1. Show that $\|\|$ is an extension of the absolute value $|\cdot|$ to $K(X)$. (Hint: The only nontrivial part is to show that $\|fg\| \geq \|f\|\|g\|$. This can be done by a careful examination of the coefficient of X^{I+J} in fg where I is given by $\|f\| = \|a_I X^I\|$ and $\|a_i X^i\| < \|f\|$ if $i < I$, and similarly for J and g.)

Until we say otherwise, assume that K is complete with respect to $|\cdot|$. Let us also assume that $a_0 \neq 0$ and $a_n \neq 0$, *i.e.*, X does not divide f and the degree of f is precisely n. To get the *Newton Polygon*, $\Pi(f)$, we consider the points in $\mathbb{R}^2$ defined by

$$P(j) = (j, \ln |a_j|) \text{ for } a_j \neq 0.$$

Then $\Pi(f)$ is the upper boundary of the convex hull of the $P(j)$. Every $P(j)$ lies on or below $\Pi(f)$. $P(0)$ and $P(n)$ are the beginning and end of $\Pi(f)$ which consists of line segments, say σ_s for $1 \leq s \leq r$ for some r. Say σ_s joins $P(m_{s-1})$ and $P(m_s)$. We get a sequence of indices

$$0 = m_0 < m_1 < \cdots < m_r = n \ .$$

The slope of σ_s is

$$\gamma_s = \frac{\ln |a_{m_s}| - \ln |a_{m_{s-1}}|}{m_s - m_{s-1}} \ ,$$

and we must have

$$\gamma_1 > \gamma_2 > \cdots > \gamma_r \ .$$

Definition 4.2.2. We say f is of *type* $(\ell_1, \gamma_1; \ell_2, \gamma_2; \ldots; \ell_r, \gamma_r)$ where $\ell_1 = m_1$ and $\ell_s = m_s - m_{s-1}$ for $s > 1$. We will usually abbreviate this by saying f is of type $(*)$.

If $r = 1$, we say f is *pure.*

So a polynomial is pure if its first and last coefficients are "bigger" than any of the rest. If a polynomial is of type (ℓ, γ) (and hence pure), then its degree must be ℓ, and $\gamma = (1/\ell) \ln |a_n/a_0|$.

Suppose f is of type $(*)$. We want to see how the Newton Polygon is related to the norm $\|\|_C$. Consider the boundary of the convex hull of the points $\{(j, \ln C^j|a_j|)\}$. Call it $\Pi_C(f)$. Let $\gamma_{C,s}$ be the corresponding slopes. Then it is an easy computation to see that

$$\gamma_{C,s} = \gamma_s + \ln C\ .$$

Fix an index s and consider the absolute value $\|\|_C$ where $C = \exp(-\gamma_s)$. Then it follows from the above equation that $\|f\|_C = \|a_j X^j\|_C$ for two fs, namely $j = m_{s-1}$ and m_s, *i.e.*, $\Pi_C(f)$ has a segment of slope zero if and only if $\ln C$ is one of the γ_s's. On the other hand, if $\ln C$ is distinct from the γ_s's, then this ($\|f\|_C = \|a_j X^j\|_C$) can only happen for precisely one value of j. Furthermore if $C = \exp(-\gamma_s)$, then it follows that

$$\|f(X) - \sum_{m_{s-1} \leq j \leq m_s} a_j X^j\|_C < \|f\|_C. \tag{4.3}$$

If one can establish inequalities of this type for particular values of C, then one gets the slopes of the lines in the Newton Polygon.

It is also easily seen that if we take $C = \exp(-\gamma)$, then f is pure of type (l, γ) if and only if

$$\|f\|_C = \|a_0\|_C = \|a_f x^f\|_C. \tag{4.4}$$

(In this case $C = |a_0/a_N|^{1/N}$ where $N = \deg f$.)

Example 4.2.3. Let us take

$$f(X) = 6 + 14X^3 - 20X^5 + 29X^8 - 23X^9 + 11X^{11} + 4X^{12} - X^{13}.$$

Its Newton Polygon is illustrated in Figure 4.1. We have

$m_0 = 0, m_1 = 3, m_2 = 5, m_3 = 8, m_4 = 9, m_5 = 11, m_6 = 12, m_7 = 13,$

while the slopes are

$\gamma_1 = 0.28, \gamma_2 = 0.18, \gamma_3 = 0.12, \gamma_4 = -0.23, \gamma_5 = -0.37, \gamma_6 = -1.01, \gamma_7 = -1.39.$

Figure 4.2 shows $\Pi_C(f)$ for $C = \exp(\gamma_2)$ and Figure 4.3 shows $\Pi_C(f)$ for $C = \exp(\gamma_3)$. In each of these cases one can see the expected flat segments.

Exercise 4.2.4. Suppose $f, g \in K[X]$ are both pure with slope γ. Then fg is also pure with slope γ.

A slightly more elaborate result is

Proposition 4.2.5. *Suppose that f is of type $(*)$ and that g is pure of type (l, γ) where $\gamma < \gamma_r$. Then fg is of type $(l_1, \gamma_1; \ldots ; l_r\gamma_r; l, \gamma)$.*

Proof. Suppose

$$g(X) = b_0 + b_1 X + \cdots + b_N X^N \ .$$

Take $C = \exp(-\gamma_s)$ for some s with $1 < s < r$. Then $\gamma_s > \gamma_r$ so $\gamma < \gamma_s$, so if we put $C_g = \exp(-\gamma)$, we get $C_g > C$. By Equation (4.4) we get $\|g\|_C = |b_0| > \|b_N\|_C$. Thus

$$\|g(X)\|_C > \|g(X) - b_0\|_C \ .$$

This and (4.3) imply

$$\|f(X)g(X) - b_0 \sum_{m_{s-1} \le j \le m_s} a_j X^j\|_C < \|fg\|_C. \qquad (4.5)$$

Similarly,

$$\|f(X)g(X) - a_n X^n g(X)\|_{C_g} < \|fg\|_{C_g} \ .$$

Now, it is not difficult to see that, Equation (4.5) says that the first r line segments of the Newton Polygon of fg must have slope γ_j for $j = 0, 1, \dots, r$ and the last equation says that the last line segment must have slope γ. □

Now we would like to state the main result that we need concerning the Newton Polygon.

Theorem 4.2.6 ("Newton")**.** *Let K be complete, and let*

$$f(X) = a_0 + a_1 X + a_2 X^2 + \cdots + a_n X^n \in K[X]$$

with $a_0 \neq 0$ and $a_n \neq 0$. Suppose N with $0 < N < n$ such that

$$\|a_N X^N\| = \|f\| \ ,$$

$$\|a_j X^j\| < \|f\| \textit{ for } j > N$$

where $\| \cdot \| = \| \cdot \|_C$ for some C. Then there are $g, h \in K[X]$ with $\deg g = N$ *and* $\deg h = n - N$ *and $f = gh$.*

The proof of this result is very much in the spirit of the proof of Hensel's Lemma.

Proof. The hypothesis implies that there is a $\Delta < 1$ such that

$$\left\| f(X) - \sum_{j=0}^{N} a_j X^j \right\| = \Delta \|f\| \ .$$

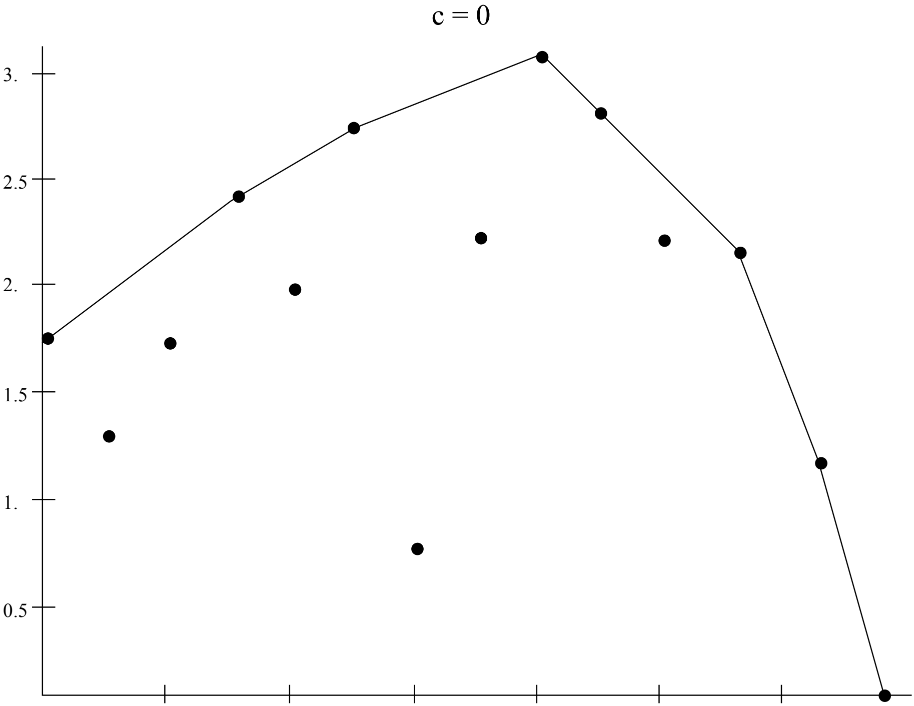
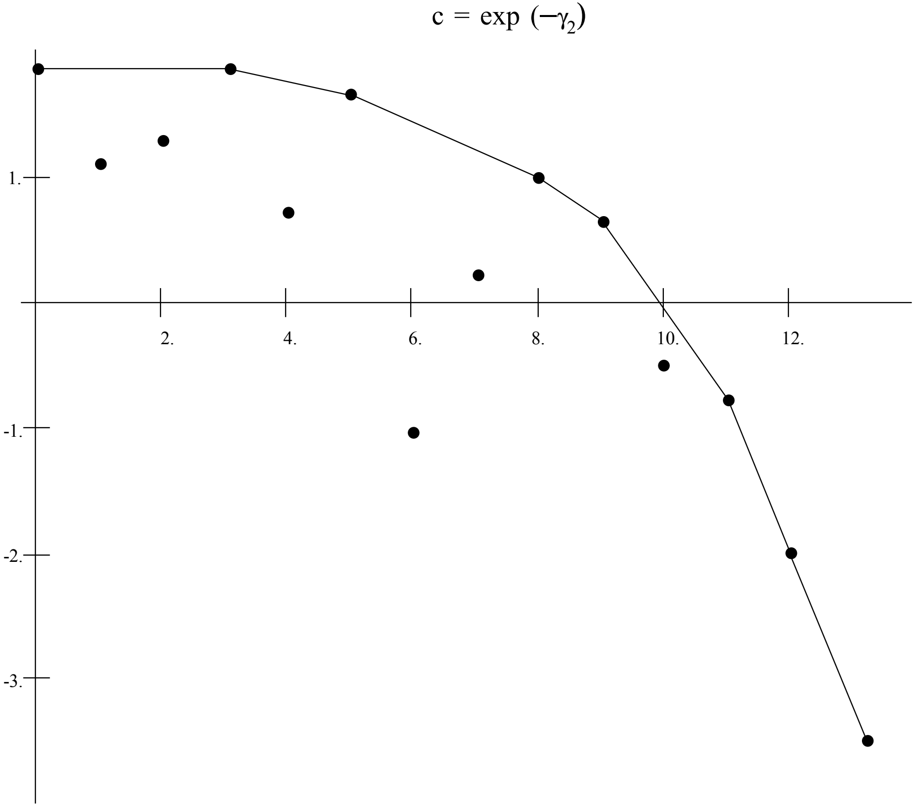


Figure 4.1



$c = \exp(-\gamma_2)$

Figure 4.2

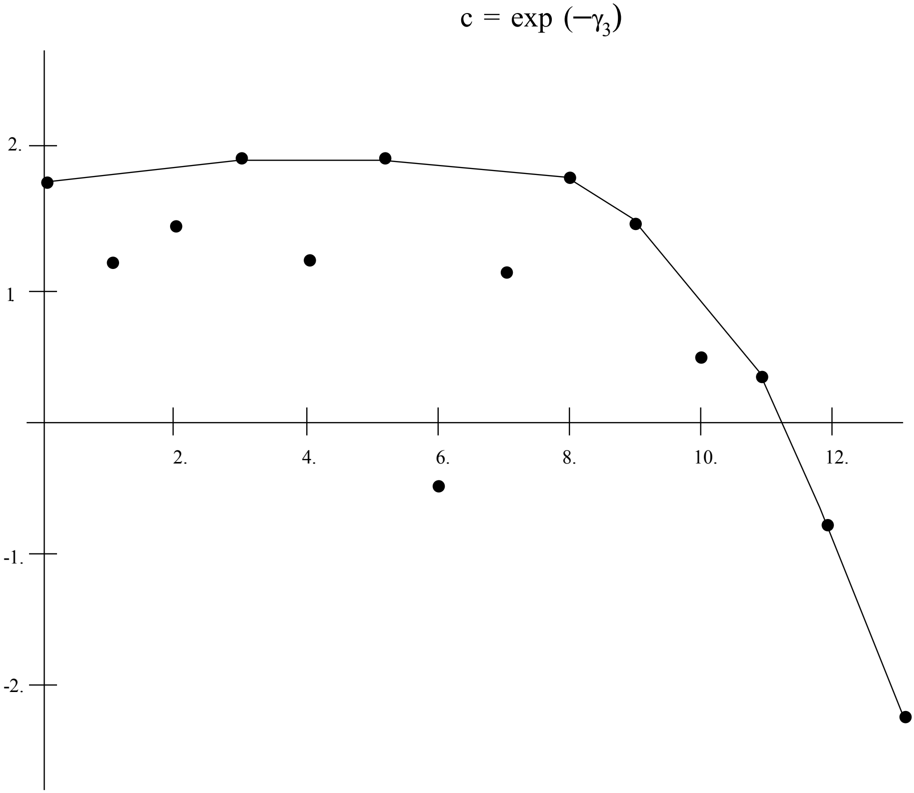


Figure 4.3

We are now going to recursively define two sequences of polynomials, $g_0, g_1, \ldots$ and $h_0, h_1, \ldots$ which converge to g and h respectively. These sequences will satisfy the following conditions:

$$\begin{aligned} \deg g_i &= N \\ \deg h_i &\leq n - N, \\ \|f - g_i\| &\leq \Delta \|f\| \ , \\ \|h_i - 1\| &\leq \Delta \ . \end{aligned}$$

If we define δ_i by

$$\|f - g_i h_i\| = \delta_i \|f\|,$$

then δ_i measures how far off $g_i h_i$ is from the true factorization $f = gh$. We want $\lim_{i \to \infty} \delta_i = 0$. In any case our conditions on g_i and h_i tell us that $\delta_i \leq \Delta$. We will arrange it so that $\delta_{i+1} \leq \Delta \delta_i$, and since $\Delta < 1$, this will insure convergence.

We get started with

$$g_0(X) = \sum_{j=0}^{N} a_j X^j,$$

$$h_0(X) = 1 \ .$$

Thus we start with $\delta_0 = \Delta$.

Now suppose we have g_i and h_i, and we want g_{i+1} and h_{i+1}. In order to conserve indexes, we write $G = g_i$ and $H = h_i$ and

$$G(X) = b_0 + b_1 X + \cdots + b_N X^N \ .$$

Since $\|f\| = \|a_N X^N\|$ and $\|f - G\| < \|f\|$, we must have $\|a_N X^N\| = \|b_N X^N\|$. Now if $\|b_i X^i\| > \|b_N X^N\|$, then we would have $\|f - G\| \geq \|(a_i - b_i)X^i\| = \|b_i X^i\| > \|a_N X^N\| = \|f\|$ which is a contradiction. Hence we have $\|G\| = \|b_N X^N\|$.

Let $D = f - GH$ and divide D by G obtaining $Q, R \in K[X]$ with $D = QG + R$ and $\deg Q \leq n - N$ and $\deg R < N$.

Claim. $\|Q\| \leq \delta_i$ and $\|R\| \leq \delta_i \|f\|$.

We have $\|D\| = \delta_i \|f\|$ and $\|G\| = \|f\|$ so if we can prove $\|Q\| \cdot \|G\| \leq \|D\|$, the first part of the claim will follow. Let $c_0, c_1, \ldots, c_{n-N}$ be the coefficients of Q. They are determined by the linear equations

$$b_N c_{n-N-j} + b_{N-1} c_{n-N-j+1} + \cdots + b_{N-j} c_{n-N} = d_{n-j} \tag{4.6}$$

where d_{n-j} is the coefficient of X^{n-j} in D, and j runs from 0 to $n - N$. We show

$$\|c_{n-N-j} X^{n-N-j}\| \cdot \|G\| \leq \|D\| \tag{4.7}$$

by induction on j. For $j = 0$, we get $b_N c_{n-N} = d_n$. Since we know that $\|G\| = C^N |b_N|$ and $\|D\| \geq C^n d_n$, the desired result follows (with equality) for $j = 0$.

Now assume we know (4.7) for $j - 1$. Then we know that all of the terms in (4.6) except for the first one have absolute value $\leq |d_{n-j}|$. If the $|b_n c_{n-N-j}| > |d_{n-j}|$, this would contradict the fact that the absolute value of the whole sum $= |d_{n-j}|$. Hence we get inequality (4.7), and the first part of the claim follows.

The second part of the claim follows easily from the first part and the equation $D = QG + R$.

Now set $g_{i+1} = G + R (= g_i + R)$ and $h_{i+1} = H + Q (= h_i + Q)$. It is trivial to see that g_{i+1} and h_{i+1} satisfy the degree condition.

The norm conditions follow from the claim, and it remains to show that $\delta_{i+1} \leq \Delta\delta_i$. We have

$$\begin{aligned} \delta_{i+1}\|f\| &= \|f - g_{i+1}h_{i+1}\| \\ &= \|f - (G+R)(H+Q)\| \\ &= \|(f - GH) - RH - QG - RQ\| \\ &= \|D - (D - R) - RH - RQ\| \\ &= \|R(H-1) + RQ\| \\ &\leq \max(\|R\| \cdot \|H - 1\|, \|R\| \cdot \|Q\|) \\ &\leq \max(\Delta\delta_i\|f\|, \delta_i^2\|f\|) \\ &\leq \Delta\delta_i\|f\| \end{aligned}$$

as desired. □

Remark. In the factorization $f = gh$ we can assume (if we so desire) that $h(0) = 1$ and that $\|h - 1\| < 1$ because we can replace h by $[h(0)]^{-1}h$.

The form in which we use the theorem is the following:

Corollary 4.2.7. *An irreducible polynomial in $K[X]$ is pure.*

Proof. Suppose f is not pure. Let

$$C = -\exp\left(\frac{1}{n}\ln\left|\frac{a_n}{a_0}\right|\right) .$$

Picking this C insures that $\|a_0\| = \|a_nX^n\|$, *i.e.*, we can think of the line between P_0 and P_n in the Newton Polygon $\Pi(f)$ as being horizontal. Since the slopes of the line segments in $\Pi(f)$ always decrease, we can find an index N such that P_N is the rightmost "largest" of the P_i's. Then it is easy to see that with this C and N, f satisfies the hypothesis of the theorem, and so cannot be irreducible. □

The next corollary is also sometimes referred to as "Newton's Theorem".

Corollary 4.2.8. *Suppose that K is complete and that $f \in K[x]$ is of type $(*)$, i.e., type $(\ell_1, \gamma_1; \ell_2, \gamma_2; \ldots; \ell_r, \gamma_r)$. Then f factors*

$$f = g_1 \cdot g_2 \cdots g_r$$

where g_s is pure of type (ℓ_s, γ_s).

Proof. Write $f = \Pi h_\lambda$ as a product of irreducible polynomials. By Corollary 4.2.7 each h_λ is pure. If more than one of the h_λ's have the same slope, then by Exercise 4.2.4 their product is pure with the same slope. Continuing in this fashion, we can write

$$f = \prod_{\lambda=1}^{M} g_\lambda$$

for some M, and say g_λ is pure of type $(q_\lambda, \delta_\lambda)$ and $\delta_1 > \delta_2 > \cdots > \delta_M$. By Proposition 4.2.5 and a little induction, the type of Πg_λ must be $(q_1, \delta_1; q_2, \delta_2; \ldots; q_M, \delta_M)$ which must be the type of f. Hence $M = r$ and $q_i = \ell_i$ and $\delta_i = \gamma_i$ for $1 \le i \le r (= M)$. □

Remark. There is another corollary of our theorem which tells us that f factors if we have a "good enough" approximate factorization. This says roughly that if $\delta = \|f - GH\|$ is less than $|R(G, H)|^2$ where $R(G, H)$ is the resultant of G and H, then $f = gh$ where $\deg g = \deg G$ and $\deg h = \deg H$. For a proof see [10], page 105.

4.3 Extensions of Valuations

Using the results of the previous section it is now relatively easy to show that we can extend a valuation (or absolute value) from a complete field to a finite extension.

Theorem 4.3.1. *Let K be a field which is complete with respect to an absolute value $|\cdot|$, and let L be an extension of K of degree n. Define a map $\|\| : L \to \mathbb{R}$ by*

$$\|x\| = |N_{L/K}x|^{1/n}$$

for $x \in L$. Then $\|\cdot\|$ is an absolute value on L which extends $|\cdot|$.

Proof. If $x \in K$, then by Proposition 3.5.3 $N_{L/K}x = x^n$ so $\|x\| = |x|$ and $\|\cdot\|$ extends $|\cdot|$. Let $x, y \in L$. By Exercise 3.5.2 $N_{L/K}xy = N_{L/K}x \cdot N_{L/K}y$ so $\|xy\| = \|x\| \cdot \|y\|$. We now want to show that $\|x+y\| \leq \max(\|x\|, \|y\|)$. Suppose $\|y\| \leq \|x\|$. Then if $z = y/x$, we have $\|z\| \leq 1$, it suffices to show that $\|z+1\| \leq 1$. Let f_z and m_z be the characteristic and minimal polynomials of z respectively. Then by Theorem 3.5.7 $f_z = m_z^r$ for some $r > 0$. Write

$$f_z(X) = X^n + a_1X^{n-1} + \cdots + a_0 \ .$$

Then $|a_0| = |\pm N_{L/K}z| \leq 1$ since $\|z\| \leq 1$. Now by Corollary 4.2.7, since m_z is irreducible, it is pure. By Exercise 4.2.4, f_z is thereby pure. Hence $|a_i| \leq |a_0| \leq 1$.

Exercise 4.3.2. Show that $N_{L/K}(1+z) = (-1)^n f_z(-1)$.

Therefore

$$\|1+z\| = |f_z(-1)|^{1/n},$$

and this is ≤ 1 since the coefficients of f_z have absolute value ≤ 1. □

We can now show that this extension of a valuation on a complete field is unique, but the proof really has little to do with the preceding material. The idea is that a finite extension L of K is a finite dimensional vector space over K, and a finite dimensional vector space over a complete field has a unique topology. Instead of using topology directly, we introduce the familiar notion of a norm on a vector space.

Definition 4.3.3. Let V be a vector space over a field K with valuation $|\cdot|$. A real valued function $\|\cdot\|$ on V is called a *norm* if the following conditions hold $\forall \vec{a}, \vec{b} \in V$ and $c \in K$:

(i) $\|\vec{a}\| \geq 0$ and $\|\vec{a}\| = 0$ if and only if $\vec{a} = 0$.

(ii) $\|\vec{a} + \vec{b}\| \leq \|\vec{a}\| + \|\vec{b}\|$.

(iii) $\|c\vec{a}\| = |c| \cdot \|\vec{a}\|$.

Remarks.

(i) This "norm" is, of course, different from the "norm" from an extension down to the base field, but this is the usual terminology. The context usually makes clear which one is intended.

(ii) A norm on V induces in the usual way a metric and thereby a topology on V. It should be clear that a norm on V lets us define cauchy sequences of points of V, and hence what it means for V to be complete.

Definition 4.3.4. Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be norms on V. Then they are said to be *equivalent* if there are $C_1, C_2 \in \mathbb{R}$ such that $\|\vec{a}\|_1 \leq C_2\|\vec{a}\|_2$ and $\|\vec{a}\|_2 \leq C_1\|\vec{a}\|_1$.

Remark. It can be shown that equivalent norms induce the same topology on V.

Here is the main result for this situation.

Theorem 4.3.5. *Suppose K is complete with respect to $|\cdot|$ and that V is a finite dimensional vector space over K. Then any two norms extending $|\cdot|$ on V are equivalent, and V is complete with respect to any such norm.*

Proof. Let $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n$ be a basis for V over K. For $\vec{a} \in K$ write

$$\vec{a} = a_1\vec{e}_1 + \cdots + a_n\vec{e}_n$$

where $a_i \in K$. We define a canonical norm on V by

$$\|\vec{a}\|_0 = \max_i |a_i| \ .$$

It is easy to see that $\|\cdot\|_0$ is a norm and that V is complete with respect to it. Hence it suffices to show that any norm on V is equivalent to $\|\cdot\|_0$.

Let $\|\cdot\|$ be any norm on V. We must establish two inequalities. One is easy.

$$\begin{aligned}\|a\| &= \|\sum_i a_i\vec{e}_i\| \\ &\le \sum_i |a_i| \cdot \|\vec{e}_i\| \\ &\le C_0\|\vec{a}\|_0\end{aligned}$$

where

$$C_0 = \sum_i \|\vec{e}_i\| \ .$$

It remains to show that there is $C \in \mathbb{R}$ such that

$$\|\vec{a}\|_0 \leq C \|a\| \quad \forall \vec{a} \in V. \tag{4.8}$$

Suppose not. We derive a contradiction by induction on the dimension n of V. Then $\forall \epsilon > 0$ there is $\vec{b} = \vec{b}(\epsilon) \in V$ such that

$$\|\vec{b}\| < \epsilon \|\vec{b}\|_0 \ .$$

By the definition of $\| \cdot \|_0$ we can assume that there is such a $\vec{b}$ with $\|\vec{b}\|_0 = |b_n|$. (Permute the $\vec{e}_i$ if necessary.) Now replace $\vec{b}$ by $b_n^{-1} \cdot \vec{b}$ so $\vec{b} = \vec{c} + \vec{e}_n$ where $\vec{c}$ is in the subspace W of V spanned by $\vec{e}_1, \ldots, \vec{e}_{n-1}$. In other words if (4.8) is false, we can find a sequence $\{\vec{c}_i\}$ of elements of W such that

$$\lim_{i \to \infty} \|\vec{c}_i + \vec{e}_n\| = 0 \ .$$

By the triangle inequality,

$$\|(\vec{c}_i + \vec{e}_n) - (\vec{c}_j + \vec{e}_n)\| \leq \|\vec{c}_i + \vec{e}_n\| + \|\vec{c}_j + \vec{e}_n\|$$

so

$$\lim_{i,j \to \infty} \|\vec{c}_i - \vec{c}_j\| = 0 \ .$$

Since the dimension of W is less than the dimension of V, by induction we get $\vec{c} \in W$ such that

$$\lim_{i \to \infty} \|\vec{c}_i - \vec{c}\| = 0 \ .$$

This implies

$$\|\vec{c} + \vec{e}_n\| = \lim_{i \to \infty} \|\vec{c}_i + \vec{e}_n\| = 0 \ .$$

But $\vec{e}_n$ is certainly not in W so $\vec{c} + \vec{e}_n \neq 0$ and we get our contradiction. Hence $\| \cdot \|_0$ and $\| \cdot \|$ are equivalent. □

Now we can easily prove the uniqueness of the extension of a valuation on a complete field.

Corollary 4.3.6. *Let K be a field which is complete with respect to an absolute value $|\cdot|$, and let L be a finite extension of K. Then there is a unique extension of $|\cdot|$ to L, and L is complete with respect to this extension.*

Proof. Theorem 4.3.1 tells us that there is an extension. If we regard L as a finite dimensional vector space over K, then it is trivial to see that the extension of the absolute value defines a norm on L. If we had two extensions of the absolute value to L, then by the theorem they would be equivalent as norms on L.

Exercise 4.3.7. Show that any two equivalent norms on L which agree on K are identical.

Proof. If N_1 and N_2 agree on K which is complete. By theorem 4.3.5, N_1 and N_2 are equivalent. Assume that there is $x \in L$ is s.t. $N_1(x) \neq N_2(x)$. Then, exchanging N_1 and N_2 if necessary, one has $N_1(x^n)/N_2(x^n) = [N_1(x)/N_2(x)]^n \to 0$ which contradicts the equivalence of the two norms.

□

The statement about the completeness of L follows immediately from the completeness of the norm on L. □

We need one more trivial fact before we can prove the required result concerning extensions in the noncomplete case.

Proposition 4.3.8. *Let K be a complete field with respect to an absolute value. Then there is a unique extension of the absolute value to $\overline{K}$, the algebraic closure of K.*

Proof. Let $a \in \overline{K}$. We know that a is algebraic over K so the extension $K(a)$ is finite over K. By Corollary 4.3.6 the absolute value on K extends uniquely to $K(a)$. For $a, b \in \overline{K}$ the extensions on $K(a)$ and $K(b)$ must also agree on $K(a) \cap K(b)$. These extensions therefore define a unique extension of the absolute value of K to $\overline{K}$. □

We can now prove Theorem 3.8.12. We restate it here in a slightly different form.

Theorem 4.3.9. *Let $L = K(x)$ be a finite separable extension of K and let $|\cdot|$ be an absolute value on K. Let $\overline{K}$ be the completion of K with respect to $|\cdot|$. Let m_x be the minimal polynomial of x. Suppose*

$$m_x = \varphi_1 \varphi_2 \cdots \varphi_r$$

is the factorization of m_x into (nonconstant) irreducibles in $\overline{K}[X]$.

Then the φ_i's are distinct. Let $L_i = \overline{K}(y_i)$ where y_i is a root of φ_i. Then there is a monomorphism

$$I_i \;:\; L = K(x) \mapsto L_i = \overline{K}(y_i)$$

which extends the monomorphism $K \mapsto \overline{K}$ under which $x \mapsto y_i$. We know that the absolute value $|\cdot|$ extends uniquely to $\overline{K}$ and thence since $\overline{K}$ is complete, uniquely to $\overline{K}(y_i) = L_j$. Using the monomorphism I_i, this defines an extension of $|\cdot|$ to L which we denote by $|\cdot|_i$. Then the absolute values $|\cdot|_1, |\cdot|_2, \ldots, |\cdot|_r$ are precisely all of the extensions of $|\cdot|$ to L. Furthermore L_i is the completion on L with respect to $|\cdot|_i$.

Proof. (As Cassels' remarks, the proof is shorter than the statement.) Let $\|\cdot\|$ be any extension of $|\cdot|$ to L, and let $\overline{L}$ be the completion of L with respect to it. Then we have $\overline{K} \subset \overline{L}$ and $x \in L \subset \overline{L}$. By Corollary 4.3.6, $\overline{K}(x)$ is complete so we must have $\overline{K}(x)$ isomorphic to $\overline{L}$ say by I. Clearly I leaves K fixed. Let $y = I(X)$, and let m_y be the minimal polynomial for y over $\overline{K}$. Since $m_x(x) = 0$, and $I(m_x) = m_x$ since $m_x \in K[X]$, we have $m_x(y) = 0$ so $m_y | m_x$. Hence m_y is one of the φ_i's as desired.

Going in the other direction, let y_i be a root of φ_i. Then $m_x(y_i) = 0$ so the extensions $K(x) = L$ and $k(y_i) \subset \overline{k}(y_i) = L_i$ are isomorphic, and we are reduced to the above situation.

It remains to show that the φ_i are distinct. If not, m_x and m_x' would have a common factor in $\overline{K}[X]$. Now we can determine this common factor by the euclidean algorithm, so it must lie in $K[X]$, but this contradicts the fact that m_x is irreducible and separable.

□

Bibliography

[1] William Fulton. *Introduction to Intersection Theory in Algebraic Geometry*, number 54 in Regional Conference Series in Mathematics, American Mathematical Society, 1984.

[2] Serge Lang. *Algebra*, Addison-Wesley, 1965.

[3] Serge Lang. *Elliptic Curves*: *Diophantine Analysis*, Springer-Verlag, 1978.

[4] Chih-Han Sah. *Abstract Algebra*, Academic Press, 1967.

[5] René Schoof. Elliptic curves over finite fields and the computation of square roots mod *p, Math. Comp.*, 44:483-494, 1985.

[6] Joseph H. Silverman. *The Arithmetic of Elliptic Curves*, number 106 in Graduate Texts in Mathematics, Springer-Verlag, 1986.

[7] Emil Artin. Galois theory, *Notre Dame Mathematical Lectures*, 2, 1957.

[8] Emil Artin. *Algebraic Numbers and Algebraic Functions*, Nelson, London, 1968.

[9] Nicolas Bourbaki. *Commutative Algebra*, Hermann, Paris, 1972.

[10] J. W. S. Cassels. *Local Fields*, Cambridge Press, 1986.

[11] William Fulton. *Algebraic Curves*, W. A. Benjamin, 1969.

[12] Shigeru Iitaka. *Algebraic Geometry*, Springer-Verlag, 1982.

[13] Burt S. Kaliski. Elliptic curves and cryptography: A pseudorandom bit generator and other tools, MIT/LSC/TR-411, 1988. Cambridge.

[14] Keith Kendig. *Elementary Algebraic Geometry*, Springer-Verlag, 1977.

[15] Paul J. McCarthy. *Algebraic Extensions of Fields*, Blaisdell, 1966.

[16] Jean-Pierre Serre. *Groupes Algébriques et Corps de Classes*, Hermann, 1959. Paris.

[17] Jean-Pierre Serre. *Local Fields*, Springer-Verlag, 1979.

[18] Edwin Weiss. *Algebraic Number Theory*, McGraw-Hill, 1963.

[19] Oscar Zariski and Pierre Samuel. *Commutative Algebra*, Van Nostrand, 1958.