4 Isogenies

In almost every branch of mathematics, when considering a category of mathematical objects with a particular structure, the maps between objects that preserve this structure (morphisms) play a crucial role. For groups and rings we have homomorphisms, for vector spaces we have linear transformations, and for topological spaces we have continuous functions. For elliptic curves (and more generally, abelian varieties), the structure-preserving maps are called *isogenies*.[1]

Remark 4.1. I have included some general background on field extensions and algebraic sets at the end of these notes (see §4.6 and §4.7) for those who have not seen this material before (or would just like a refresher).

4.1 Morphisms of projective curves

As abelian varieties, elliptic curves have both an algebraic structure (as an abelian group), and a geometric structure (as a smooth projective curve). We are all familiar with morphisms of groups (these are group homomorphisms), but we have not formally defined a morphism of projective curves. To do so we need to define a few notions from algebraic geometry. Since algebraic geometry is not a prerequisite for this course, we will take a brief detour to define the terms we need.

To keep things as simple and concrete as possible, we will focus on plane projective curves with a few remarks along the way about how to generalize these definitions for those who are interested (those who are not can safely ignore the remarks). As usual, we use $\bar{k}$ to denote a fixed algebraic closure of our base field k that contains any and all algebraic extensions of k that we may consider (see §4.6 for more on algebraic closures).

Definition 4.2. Let C/k be a plane projective curve $f(x,y,z) = 0$ with f a nonconstant homogeneous polynomial in $k[x,y,z]$ that is irreducible in $\bar{k}[x,y,z]$. The *function field* $k(C)$ is the set of equivalence classes of rational functions g/h such that:

(i) g and h are homogeneous polynomials in $k[x,y,z]$ of the same degree;
(ii) h is not divisible by f, equivalently, h is not an element of the ideal (f);
(iii) g_1/h_1 and g_2/h_2 are considered equivalent whenever $g_1h_2 - g_2h_1 \in (f)$.

If L is any algebraic extension of k (including $L = \bar{k}$), the function field $L(C)$ is similarly defined with $g, h \in L[x,y,z]$.

Remark 4.3. The function field $k(X)$ of an irreducible projective variety X/k given by homogeneous polynomials $f_1, \dots, f_m \in k[x_0, \dots, x_n]$ is defined similarly: just replace the homogeneous ideal (f) with the homogeneous ideal $(f_1, \dots, f_m)$ (homogeneous ideal means an ideal of $k[x_0, \dots, x_n]$ generated by homogeneous polynomials).

Remark 4.4. Be sure not to confuse the notation $k(C)$ with $C(k)$; the latter denotes the set of k-rational points on C, not its function field.

[1] The word *isogeny* literally means "equal origins". It comes from biology, where the terms *isogenous*, *isogenic*, and *isogenetic* refer to different tissues derived from the same progenitor cell. The prefix "iso" means equal and the root "gene" means origin (as in the word *genesis*).

We claim that $k(C)$ is a ring under addition and multiplication of rational functions. To see this, first note that if $h_1, h_2 \notin (f)$ then $h_1 h_2 \notin (f)$ because f is irreducible and $k[x,y,z]$ is a unique factorization domain (in particular, (f) is a prime ideal). Thus for any $g_1/h_1, g_2/h_2 \in k(C)$ we have

$$\frac{g_1}{h_1} + \frac{g_2}{h_2} = \frac{g_1 h_2 + g_2 h_1}{h_1 h_2} \in k(C) \qquad \text{and} \qquad \frac{g_1}{h_1} \cdot \frac{g_2}{h_2} = \frac{g_1 g_2}{h_1 h_2} \in k(C).$$

We can compute the inverse of g/h as h/g except when $g \in (f)$, but in this case g/h is equivalent to $0/1 = 0$, since $g \cdot 1 - 0 \cdot h = g \in (f)$; thus every nonzero element of $k(C)$ is invertible, hence the ring $k(C)$ is a field.

Remark 4.5. The field $k(C)$ contains k as a subfield (take g and h with degree 0), but it is not an algebraic extension of k, it is transcendental. Indeed, it has transcendence degree 1, consistent with the fact that C is a projective variety of dimension 1 (this is one way to define the dimension of an algebraic variety). See §4.6 for more on transcendental field extensions.

The fact that g and h have the same degree allows us to meaningfully assign a value to the function g/h at a projective point $P = (x_0 : y_0 : z_0)$ on C, so long as $h(P) \neq 0$, since

(a) we get the same result for any projectively equivalent $P = (\lambda x_0 : \lambda y_0 : \lambda z_0)$ with $\lambda \in k^\times$, because g and h are homogeneous of the same degree (say d):

$$\frac{g(\lambda x, \lambda y, \lambda z)}{h(\lambda x, \lambda y, \lambda z)} = \frac{\lambda^d g(x,y,z)}{\lambda^d h(x,y,z)} = \frac{g(x,y,z)}{h(x,y,z)}.$$

(b) if g_1/h_1 and g_2/h_2 are equivalent and $h_1(P), h_2(P) \neq 0$, then $g_1(P)h_2(P) - g_2(P)h_1(P)$ is a multiple of $f(P) = 0$, so $(g_1/h_1)(P) = (g_2/h_2)(P)$.

Thus assuming the denominators involved are all nonzero, for $\alpha \in k(C)$ the value of $\alpha(P)$ does not depend on how we choose to represent either α or P. If $\alpha = g_1/h_1$ with $h_1(P) = 0$, it may happen that g_1/h_1 is equivalent to some g_2/h_2 with $h_2(P) \neq 0$. This is a slightly subtle point. It may not be immediately obvious whether or not such a g_2/h_2 exists, since it depends on equivalence modulo f; in general there may be no canonical way to write g/h in "lowest terms", because the ring $k[x,y,z]/(f)$ is typically *not* a unique factorization domain.

Example 4.6. Suppose C/k is defined by $f(x,y,z) = zy^2 - x^3 - z^2x = 0$, and consider the point $P = (0:0:1) \in C(k)$. We can't evaluate $\alpha = 3xz/y^2 \in k(C)$ at P as written since its denominator vanishes at P, but we can use the equivalence relation in $k(C)$ to write

$$\alpha = \frac{3xz}{y^2} = \frac{3xz^2}{x^3 + z^2x} = \frac{3z^2}{x^2 + z^2},$$

and we then see that $\alpha(P) = 3$.

Definition 4.7. Let C/k be a projective curve with $\alpha \in k(C)$. We say that α is *defined* (or *regular*) at a point $P \in C(\bar{k})$ if α can be represented as g/h for some $g, h \in k[x,y,z]$ with $h(P) \neq 0$.

Remark 4.8. If C is the projective closure of an affine curve $f(x,y) = 0$, one can equivalently define $k(C)$ as the fraction field of $k[x,y]/(f)$; this ring is known as the *coordinate ring* of C, denoted $k[C]$, and it is an integral domain provided that (f) is a prime ideal (which holds in our setting because we assume f is irreducible). In this case one needs to homogenize rational functions $r(x,y) = g(x,y)/h(x,y)$ in order to view them as functions defined on projective space. This is done by introducing powers of z so that the numerator and denominator are homogeneous polynomials of the same degree. The same remark applies to (irreducible) varieties of higher dimension.

Recall that for any field F (including $F = k(C)$), we use $\mathbb{P}^2(F)$ to denote the set of projective triples $(x : y : z)$, with $x, y, z \in F$ not all zero, modulo the equivalence relation $(x : y : z) \sim (\lambda x : \lambda y : \lambda z)$ for $\lambda \in F^\times$.

Definition 4.9. Let C_1 and C_2 be plane projective curves defined over k. A *rational map* $\phi\colon C_1 \to C_2$ is a projective triple $(\phi_x : \phi_y : \phi_z) \in \mathbb{P}^2(k(C_1))$, such that for every $P \in C_1(\bar{k})$ where ϕ_x, ϕ_y, ϕ_z are defined and not all zero, the projective point $(\phi_x(P) : \phi_y(P) : \phi_z(P))$ lies in $C_2(\bar{k})$. The map ϕ is *defined* (or *regular*) at P if there exists $\lambda \in k(C_1)^\times$ such that $\lambda\phi_x, \lambda\phi_y, \lambda\phi_z$ are all defined at P and not all zero at P.

Remark 4.10. This definition generalizes to projective varieties in $\mathbb{P}^n$ in the obvious way.

We should note that a rational map is not simply a function from $C_1(k)$ to $C_2(k)$ defined by rational functions, for two reasons. First, it might not be defined everywhere (although for smooth projective curves this does not happen, by Theorem 4.15 below). Second, it is required to map $C_1(\bar{k})$ to $C_2(\bar{k})$, which does not automatically hold for every rational map the carries $C_1(k)$ to $C_2(k)$; indeed, in general $C_1(k)$ could be the empty set (if C_1 is an elliptic curve then $C_1(k)$ is nonempty, but it could contain just a single point).

Remark 4.11. This is a general feature of classical algebraic geometry. In order for the definitions to work properly, one must consider the situation over an algebraic closure. An alternative and much more general approach is to use *schemes*, but this requires more material than we have time to develop in this course (take 18.725/6 to learn about schemes).

It is important to remember that a rational map $\phi = (\phi_x : \phi_y : \phi_z)$ is defined only up to scalar equivalence by functions in $k(C)^\times$. There may be points $P \in C_1(\bar{k})$ where one of $\phi_x(P), \phi_y(P), \phi_z(P)$ is not defined or all three are zero, but it may still possible to evaluate $\phi(P)$ after rescaling by $\lambda \in k(C)^\times$; we will see an example of this shortly.

The value of $\phi(P)$ is unchanged if we clear denominators in $(\phi_x : \phi_y : \phi_z)$ by multiplying through by an appropriate homogeneous polynomial (note: this is not the same as rescaling by an element of $\lambda \in k(C)^\times$). This yields a triple $(\psi_x : \psi_y : \psi_z)$ of homogeneous polynomials of equal degree that we view as a representing any of the three equivalent rational maps

$$(\psi_x/\psi_z : \psi_y/\psi_z : 1), \qquad (\psi_x/\psi_y : 1 : \psi_z/\psi_y), \qquad (1 : \psi_y/\psi_x : \psi_z/\psi_x),$$

all of which are equivalent to ϕ. We then have $\phi(P) = (\psi_x(P) : \psi_y(P) : \psi_z(P))$ whenever any of ψ_x, ψ_y, ψ_z is nonzero at P. Of course it can still happen that ψ_x, ψ_y, ψ_z all vanish at P, in which case we might need to look for an equivalent tuple of homogeneous polynomials that represents ϕ. The tuples $(\psi_x : \psi_y : \psi_z)$ and $(\psi'_x : \psi'_y : \psi'_z)$ represent the same rational map whenever the polynomials $\psi_x\psi'_y - \psi'_x\psi_y$ and $\psi_x\psi'_z - \psi'_x\psi_z$ and $\psi_y\psi'_z - \psi'_y\psi_z$ all lie in the ideal (f_1) defining C_1.

This defines an equivalence relation on set of triples $(\psi_x : \psi_y : \psi_z)$ of nonzero homogeneous polynomials of the same degree that satisfy $f_2(\psi_x, \psi_y, \psi_2) \in (f_1)$, where (f_2) is the ideal defining C_2. Each equivalence class corresponds to a rational map $C_1 \to C_2$ and every rational map has a corresponding equivalence class.

Remark 4.12. This set of equivalence classes of tuples defining rational maps $\psi\colon V_1 \to V_2$ of projective varieties also generalizes: replace (f_1) with the homogeneous ideal I_1 defining V_1 and require $f_2(\psi) \in I_1$ for every generator f_2 of the homogeneous ideal I_2 defining V_2.

This leads to the following equivalent definition of a rational map.

Definition 4.13. Let C_1 and C_2 be plane projective curves over k defined by $f_1(x, y, z) = 0$ and $f_2(x, y, z) = 0$, respectively. A *rational map* $\psi\colon C_1 \to C_2$ is an equivalence class of triples $(\psi_x : \psi_y : \psi_z)$ of homogeneous polynomials in $k[x, y, z]$ of the same degree, not all of which lie in (f_1), such that $f_2(\psi_x, \psi_y, \psi_z) \in (f_1)$. Triples $(\psi_x : \psi_y : \psi_z)$ and $(\psi'_x : \psi'_y : \psi'_z)$ are equivalent whenever $\psi_x\psi'_y - \psi'_x\psi_y$ and $\psi_x\psi'_z - \psi'_x\psi_z$ and $\psi_y\psi'_z - \psi'_y\psi_z$ all lie in (f_1).

The rational map ϕ is *defined* at $P \in C_1(\bar{k})$ if any of $\psi_x(P), \psi_y(P), \psi_z(P)$ is nonzero, in which case $(\psi_x(P) : \psi_y(P) : \psi_z(P) \in C_2(\bar{k}))$.

The equivalence of Definitions 4.9 and 4.13 follows from Corollary 4.52 (see §4.7).

Definition 4.14. A rational map that is defined everywhere is called a *morphism*.

For elliptic curves, distinguishing rational maps from morphisms is unnecessary; every rational map between elliptic curves is a morphism. More generally, we have the following.

Theorem 4.15. *If C_1 is a smooth projective curve then every rational map from C_1 to a projective curve C_2 is a morphism.*

The proof of this theorem is straightforward (see [6, II.2.1]), but requires a bit of commutative algebra that is outside the scope of this course.[2]

Remark 4.16. Theorem 4.15 is specific to smooth curves; it is not true more generally.

Two projective curves C_1 and C_2 are *isomorphic* if they are related by an invertible morphism ϕ; this means that there is a morphism ϕ^{-1} such that $\phi^{-1} \circ \phi$ and $\phi \circ \phi^{-1}$ are the identity maps on $C_1(\bar{k})$ and $C_2(\bar{k})$, respectively. An isomorphism $\phi\colon C_1 \to C_2$ is necessarily a morphism that defines a bijection from $C_1(\bar{k})$ to $C_2(\bar{k})$, but the converse is not true, in general, because the inverse map of sets from $C_2(\bar{k})$ to $C_1(\bar{k})$ might not be a morphism (because it can't be defined by rational functions); we will see an example of this shortly.

Before leaving the topic of morphisms of curves, we note one more useful fact.

Theorem 4.17. *A morphism of projective curves is either surjective or constant.*

This theorem is a consequence of the fact that projective varieties are *complete* (or *proper*), which implies that the image of a morphism of projective varieties is itself a projective variety. This is a standard result that is proved in most introductory algebraic geometry textbooks, see [2, II.4.9], for example. In the case of projective curves the image of a morphism $\phi\colon C_1 \to C_2$ of curves either has dimension 1, in which case ϕ is surjective (our curves are irreducible, by definition, and therefore cannot properly contain another curve), or dimension 0, in which case the image is a single point and ϕ is constant.

[2]The key point is that the coordinate ring of a smooth curve is a Dedekind domain. Thus its localization at every point P is a DVR, and after choosing a uniformizer we can rescale any rational map ϕ by a suitable λ (which will typically vary with P) so that all the components of ϕ have non-negative valuation at P and at least one has valuation zero and is therefore nonvanishing at P.

4.2 Isogenies of elliptic curves

We can now define the structure-preserving maps between elliptic curves that will play a key role in this course.

Definition 4.18. An *isogeny* $\phi\colon E_1 \to E_2$ of elliptic curves defined over k is a surjective morphism of curves that induces a group homomorphism $E_1(\bar{k}) \to E_2(\bar{k})$. The elliptic curves E_1 and E_2 are then said to be *isogenous*.

Remark 4.19. Unless otherwise stated, we assume that the isogeny ϕ is itself defined over k (meaning that it can be represented by a rational map whose coefficients lie in k). In general, if L/k is an algebraic extension, we say that two elliptic curves defined over k are "isogenous over L" if they are related by an isogeny that is defined over L. Strictly speaking, in this situation we are really referring to the "base change" of the elliptic curves to L (same equations, different field of definition), but we won't be pedantic about this.

This definition is stronger than is actually necessary, for three reasons. First, any morphism of abelian varieties that preserves the identity element (the distinguished point that is the zero element of the group) induces a group homomorphism; we won't bother to prove this (see [6, Theorem III.4.8] for a proof), since for all the isogenies we are interested in it will be obvious that they are group homomorphisms. Second, by Theorem 4.17, any non-constant morphism of curves is surjective, and third, by Theorem 4.15, a rational map whose domain is a smooth projective curve is automatically a morphism. This leads to the following equivalent definition which is commonly used.

Definition 4.20. An *isogeny* $\phi\colon E_1 \to E_2$ of elliptic curves defined over k is a non-constant rational map that sends the distinguished point of E_1 to the distinguished point of E_2.

Warning 4.21. Under our definitions the zero morphism, which maps every point on E_1 to the zero point of E_2, is *not* an isogeny. This follows the standard convention for general abelian varieties which requires isogenies to preserve dimension (so they must be surjective and have finite kernel). In the case of elliptic curves this convention is not always followed (notably, Silverman [6, III.4] includes the zero morphism in his definition of an isogeny), but it simplifies the statement of many theorems and is consistent with the more general usage you may see in later courses, so we will use it (we will still have occasion to refer to the zero morphism, we just won't call it an isogeny).

Definition 4.22. Elliptic curves E_1 and E_2 defined over a field k are *isomorphic* if there exist isogenies $\phi_1\colon E_1 \to E_2$ and $\phi_2\colon E_2 \to E_1$ whose composition is the identity; the isogenies ϕ_1 and ϕ_2 are then *isomorphisms*.

Definition 4.23. A morphism from an elliptic curve E/k to itself that fixes the distinguished point is called an *endomorphism*. An endomorphism that is also an isomorphism is an *automorphism*.

Except for the zero morphism, every endomorphism is an isogeny. As we shall see in the next lecture, the endomorphisms of an elliptic curve have a natural ring structure.

4.3 Examples of isogenies

We now give three examples of isogenies that are endomorphisms of an elliptic curve E/k defined by a short Weierstrass equation $y^2 = x^3 + Ax + b$ (we assume $\mathrm{char}(k) \neq 2, 3$).

4.3.1 The negation map

In projective coordinates the map $P \mapsto -P$ is given by

$$(x : y : z) \mapsto (x : -y : z),$$

which is evidently a rational map. It is defined at every projective point, and in particular, at every $P \in E(\bar{k})$, so it is a morphism (as it must be, since it is a rational map defined on a smooth curve). It fixes $0 = (0 : 1 : 0)$ and is not constant, thus it is an isogeny. It is also an endomorphism, since it is a morphism from E to E that fixes 0, and moreover an isomorphism (it is its own inverse), and therefore an automorphism.

4.3.2 The multiplication-by-2 map

Let E/k be the elliptic curve defined by $y^2 = x^3 + Ax + B$, and let $\phi\colon E \to E$ be defined by $P \mapsto 2P$. This is obviously a non-trivial group homomorphism (at least over $\bar{k}$), and we will now show that it is a morphism of projective curves. Recall that the formula for doubling an affine point $P = (x, y)$ on E is given by the rational functions

$$\phi_x(x, y) = m(x, y)^2 - 2x = \frac{(3x^2 + A)^2 - 8xy^2}{4y^2},$$
$$\phi_y(x, y) = m(x, y)(x - \phi_x(x, y)) - y = \frac{12xy^2(3x^2 + A) - (3x^2 + A)^3 - 8y^4}{8y^3},$$

where $m(x, y) := (3x^2 + A)/(2y)$ is the slope of the tangent line at P. Homogenizing these and clearing denominators yields the rational map $\phi := (\psi_x/\psi_z : \psi_y/\psi_z : 1)$, where

$$\psi_x(x, y, z) = 2yz\big((3x^2 + Az^2)^2 - 8xy^2z\big),$$
$$\psi_y(x, y, z) = 12xy^2z(3x^2 + Az^2) - (3x^2 + Az^2)^3 - 8y^4z^2,$$
$$\psi_z(x, y, z) = 8y^3z^3.$$

If $y = 0$ then $3x^2 + Az^2 \neq 0$ (because $y^2z = x^3 + Axz^2 + Bz^3$ is non-singular), and it follows that the only point in $E(\bar{k})$ where ψ_x, ψ_y, ψ_z simultaneously vanish is the point $0 = (0 : 1 : 0)$. As a rational map of smooth projective curves, we know that ϕ is a morphism, hence defined everywhere, so there must be an alternative representation of ϕ that we can evaluate at the point 0. Now in fact we know *a priori* that $\phi(0)$ must be 0, since $2 \cdot 0 = 0$ but let's verify this explicitly.

In projective coordinates the curve equation is $f(x, y, z) := y^2z - x^3 - Axz^2 - Bz^3 = 0$. We are free to add any multiple of f in $k[x, y, z]$ of the correct degree (in this case 6) to any of ψ_x, ψ_y, ψ_z without changing the rational function ϕ they define. Let us replace ψ_x with $\psi_x + 18xyzf$ and ψ_y with $\psi_y + (27f - 18y^2z)f$, and remove the common factor z^2 to obtain

$$\psi_x(x, y, z) = 2y\big(xy^2 - 9Bxz^2 + A^2z^3 - 3Ax^2z\big),$$
$$\psi_y(x, y, z) = y^4 - 12y^2z(2Ax + 3Bz) - A^3z^4$$
$$\qquad + 27Bz(2x^3 + 2Axz^2 + Bz^3) + 9Ax^2(3x^2 + 2Az^2),$$
$$\psi_z(x, y, z) = 8y^3z.$$

This is another representation of the rational map ϕ, and we can use this representation of ϕ to evaluate $\phi(0) = (\psi_x(0, 1, 0) : \psi_y(0, 1, 0) : \psi_z(0, 1, 0)) = (0 : 1 : 0) = 0$, as expected.

Having seen how messy things can get even with the relatively simple isogeny $P \mapsto 2P$, in the future we will be happy to omit such verifications and rely on the fact that if we have a rational map that we know represents an isogeny ϕ, then $\phi(0) = 0$ must hold. For elliptic curves in Weierstrass form, this means we only have to worry about evaluating isogenies at affine points, which allows us to simplify the equations by fixing $z = 1$.

4.3.3 The Frobenius endomorphism

Let $\mathbb{F}_p$ be a finite field of prime order p. The *Frobenius automorphism* $\pi\colon \overline{\mathbb{F}}_p \to \overline{\mathbb{F}}_p$ is the map $x \mapsto x^p$. It is easy to check that π is a field automorphism: $0^p = 0$, $1^p = 1$, $(-a)^p = -a^p$, $(a^{-1})^p = (a^p)^{-1}$, $(ab)^p = a^p b^p$, and $(a+b)^p = \sum \binom{p}{k} a^k b^{p-k} = a^p + b^p$. If $f(x_1, \ldots, x_k)$ is any rational function with coefficients in $\mathbb{F}_p$, then

$$f(x_1, \ldots, x_k)^p = f(x_1^p, \ldots, x_k^p),$$

since the coefficients of f are all fixed by π, which acts trivially on $\mathbb{F}_p$.

Every power π^n of π is also an automorphism of $\overline{\mathbb{F}}_p$; the fixed field of π^n is the finite field $\mathbb{F}_{p^n}$ with p^n elements. For a finite field $\mathbb{F}_q = \mathbb{F}_{p^n}$ the map $x \mapsto x^q$ is called the *q-power Frobenius map*, which we may denote by π_q.

Definition 4.24. Let E be an elliptic curve over a finite field $\mathbb{F}_q$. The *Frobenius endomorphism* of E is the map $\pi_E\colon (x : y : z) \mapsto (x^q : y^q : z^q)$.

To see that this defines a morphism from E to E, for any point $P = (x, y, z) \in E(\overline{\mathbb{F}}_q)$, if we raise both sides of the curve equation

$$y^2 z = x^3 + Axz^2 + Bz^3$$

to the qth power, we get

$$\begin{aligned}(y^2 z)^q &= (x^3 + Axz^2 + Bz^3)^q \\ (y^q)^2 z^q &= (x^q)^3 + Ax^q (z^q)^2 + B(z^q)^3,\end{aligned}$$

thus $(x^q : y^q : z^q) \in E(\overline{\mathbb{F}}_q)$; we have $A^q = A$ and $B^q = B$ because $A, B \in \mathbb{F}_q$. Note that when $q \neq p$ applying the p-power Frobenius yields a point on the elliptic curve $y^2 = x^3 + A^p x + B^p$, and unless $A, B \in \mathbb{F}_p$ this won't be the same curve as E (or even isomorphic to E, in general).

To see that π_E is also a group homomorphism, note that the group law on E is defined by rational functions whose coefficients lie in $\mathbb{F}_q$; these coefficients are invariant under the q-power map, so $\pi_E(P + Q) = \pi_E(P) + \pi_E(Q)$ for all $P, Q \in E(\overline{\mathbb{F}}_q)$.

These facts hold regardless of the equation used to define E and the formulas for the group law, including curves defined by a general Weierstrass equation (which is needed in characteristic 2 and 3).

Remark 4.25. The Frobenius endomorphism induces a group isomorphism from $E(\overline{\mathbb{F}}_q)$ to $E(\overline{\mathbb{F}}_q)$, since over the algebraic closure we can take qth roots of coordinates of points, and doing so still fixes elements of $\mathbb{F}_q$ (in other words, the inverse of π_q in $\mathrm{Gal}(\overline{\mathbb{F}}_q/\mathbb{F}_q)$ also commutes with the group operation). But as an *isogeny* the Frobenius endomorphism is *not* an isomorphism because there is no rational map from $E \to E$ that acts as its inverse (why this is so will become obvious in later lectures).

4.4 A standard form for isogenies

To facilitate our work with isogenies, it will be convenient to put them in a standard form. In order to do so we will assume throughout that we are working with elliptic curves of the form $y^2 = f(x)$, and when it is convenient we will further assume $f(x) = x^3 + Ax + B$ so that our curves are in short Weierstrass form. Implicit in this assumption is that our elliptic curves are defined over a field k whose characteristic is not 2, and when we assume $f(x) = x^3 + Ax + B$ we eliminate some elliptic curves in characteristic 3.

Lemma 4.26. *Let $E_1\colon y^2 = f_1(x)$ and $E_2\colon y^2 = f_2(x)$ be elliptic curves over k, and let $\alpha\colon E_1 \to E_2$ be an isogeny. Then α can be defined by an affine rational map of the form*

$$\alpha(x,y) = \left(\frac{u(x)}{v(x)}, \frac{s(x)}{t(x)}y\right),$$

where $u, v, s, t \in k[x]$ are polynomials in x with $u \perp v$ and $s \perp t$.

The notation $u \perp v$ indicates that the polynomials u and v are coprime ($\gcd(u,v) = 1$).

Proof. Suppose α is defined by the rational map $(\alpha_x : \alpha_y : \alpha_z)$. Then for any affine point $(x : y : 1) \in E_1(\bar{k})$ we can write

$$\alpha(x,y) = \Big(r_1(x,y), r_2(x,y)\Big),$$

with $r_1(x,y) := \alpha_x(x,y,1)/\alpha_z(x,y,1)$ and $r_2(x,y) := \alpha_y(x,y,1)/\alpha_z(x,y,1)$. By repeatedly using the curve equation $y^2 = f_1(x)$ for E_1 to replace y^2 with $f_1(x)$, we can assume that both r_1 and r_2 have degree at most 1 in y. We then have

$$r_1(x,y) = \frac{p_1(x) + p_2(x)y}{p_3(x) + p_4(x)y},$$

for some $p_1, p_2, p_3, p_4 \in k[x]$. We now multiply the numerator and denominator of $r_1(x,y)$ by $p_3(x) - p_4(x)y$, and use the curve equation for E_1 to replace the y^2 in the denominator with $f_1(x)$, putting r_1 in the form

$$r_1(x,y) = \frac{q_1(x) + q_2(x)y}{q_3(x)},$$

for some $q_1, q_2, q_3 \in k[x]$.

We now use the fact that α is a group homomorphism and must therefore satisfy $\alpha(-P) = -\alpha(P)$ for any $P \in E_1(\bar{k})$. Recall that the inverse of an affine point (x,y) on a curve in short Weierstrass form is $(x,-y)$. Thus $\alpha(x,-y) = -\alpha(x,y)$ and we have

$$\Big(r_1(x,-y), r_2(x,-y)\Big) = \Big(r_1(x,y), -r_2(x,y)\Big)$$

Thus $r_1(x,y) = r_1(x,-y)$, and this implies that q_2 is the zero polynomial. After eliminating any common factors from q_1 and q_3, we obtain $r_1(x,y) = \frac{u(x)}{v(x)}$ for some $u, v \in k[x]$ with $u \perp v$, as desired. The argument for $r_2(x,y)$ is similar, except now we use $r_2(x,-y) = -r_2(x,y)$ to show that q_1 must be zero, yielding $r_2(x,y) = \frac{s(x)}{t(x)}y$ for some $s, t \in k[x]$ with $s \perp t$. □

We shall refer to the expression $\alpha(x,y) = \left(\frac{u(x)}{v(x)}, \frac{s(x)}{t(x)}y\right)$ given by Lemma 4.26 as the *standard form* of an isogeny $\alpha\colon E_1 \to E_2$. This expression represents the isogeny $\alpha\colon E_1 \to E_2$ as an affine point in $\mathbb{P}^2(k(E_1))$ whose z-coordinate α_z is the constant function 1, which means that α_x and α_y are uniquely determined as elements of $k(E_1)$. The rational functions representing α_x and α_y as elements of $k(E_1)$ are uniquely determined if we also require that

- the numerator and denominator of α_x should be homogeneous coprime polynomials in $k[x,z]$ of the same degree.
- the numerator and denominator of α_y should be homogeneous coprime polynomials in $k[x,y,z]$ of the same degree, such that the numerator has the form $g(x,z)y$.

If we additionally require that u and s be monic, the polynomials $u,v,s,t \in k[x]$ that appear in the standard form of $\alpha(x,y)$ are uniquely determined by α.

Lemma 4.27. *Let $E_1\colon y^2 = f_1(x)$ and $E_2\colon y^2 = f_2(x)$ be elliptic curves over k and let $\alpha(x,y) = \left(\frac{u(x)}{v(x)}, \frac{s(x)}{t(x)}y\right)$ be an isogeny from E_1 to E_2 in standard form. Then v^3 divides t^2 and t^2 divides $v^3 f_1$. Moreover, $v(x)$ and $t(x)$ have the same set of roots in $\bar{k}$.*

Proof. Substituting $\left(\frac{u}{v}, \frac{s}{t}y\right)$ for (x,y) in the equation for E_2 gives $((s/t)y)^2 = f_2(u/v)$, and using the equation for E_1 to replace y^2 with $f_1(x)$ yields

$$(s/t)^2 f_1 = f_2(u/v)$$

as an identity involving polynomials $f_1, f_2, s, t, u, v \in k[x]$. If we put $w = v^3 f_2(u/v) \in k[x]$ and clear denominators we obtain

$$v^3 s^2 f_1 = t^2 w. \tag{1}$$

Note that $u \perp v$ implies $v \perp w$, since any common factor of v and w must divide u. It follows that $v^3 | t^2$ and $t^2 | v^3 f_1$. This implies that v and t have the same roots in $\bar{k}$: every root of v is clearly a root of t (since $v^3 | t^2$), and every root x_0 of t is a double root of $t^2 | v^3 f_1$, and since f_1 has no double roots (because E_1 is not singular), x_0 must be a root of v (and possibly also a root of f_1). □

Corollary 4.28. *Let $\alpha(x,y) = \left(\frac{u(x)}{v(x)}, \frac{s(x)}{t(x)}y\right)$ be an isogeny $E_1 \to E_2$ in standard form. The affine points $(x_0 : y_0 : 1) \in E_1(\bar{k})$ in the kernel of α are precisely those for which $v(x_0) = 0$.*

Proof. If $v(x_0) \neq 0$, then $t(x_0) \neq 0$, and $\alpha(x_0, y_0) = \left(\frac{u(x_0)}{v(x_0)}, \frac{s(x_0)}{t(x_0)}y\right)$ is an affine point and therefore not 0 (the point at infinity), hence not in the kernel of α.

By homogenizing and putting α into projective form, we can write α as

$$\alpha = (ut : vsy : vt),$$

where ut, vsy, and vt are now homogeneous polynomials of equal degree ($s,t,u,v \in k[x,z]$).

Suppose $y_0 \neq 0$. By the previous lemma, if $v(x_0,1) = 0$, then $t(x_0,1) = 0$, and since $v^3 | t^2$, the multiplicity of $(x_0,1)$ as a root of t is strictly greater than its multiplicity as a root of v. This implies that, working over $\bar{k}$, we can renormalize α by dividing by a suitable power of $x - x_0 z$ so that α_y does not vanish at $(x_0 : y_0 : 1)$ but α_x and α_z both do. Then $\alpha(x_0 : y_0 : 1) = (0 : 1 : 0) = 0$, and $(x_0 : y_0 : 1)$ lies in the kernel of α as claimed.

If $y_0 = 0$, then x_0 is a root of the cubic $f(x)$ in the equation $y^2 = f_1(x)$ for E_1, and it is not a double root, since E_1 is not singular. In this case we renormalize α by multiplying

by yz and then replacing y^2z with $f_1(x,z)$. Because $(x_0,1)$ only has multiplicity 1 as a root of $f_1(x,z)$, its multiplicity as a root of vf_1 is no greater than its multiplicity as a root of t (here again we use $v^3|t^2$), and we can again renormalize α by dividing by a suitable power of $x - x_0z$ so that α_y does not vanish at $(x_0 : y_0 : 1)$, but α_x and α_z do (since they are now both divisible by $y_0 = 0$). Thus $(x_0 : y_0 : 1)$ is again in the kernel of α. □

The corollary implies that if we have an isogeny $\alpha\colon E_1 \to E_2$ in standard form, we know exactly what to do whenever we get a zero in the denominator when we try to compute $\alpha(P)$: we must have $\alpha(P) = 0$. This allows us to avoid in all cases the messy process that we went through earlier with the multiplication-by-2 map. We also obtain the following.

Corollary 4.29. *Let $\alpha\colon E_1 \to E_2$ be an isogeny of elliptic curves defined over a field k. The kernel of α is a finite subgroup of $E_1(\bar{k})$*

This corollary is true in general, but we will prove it under the assumption that we can put the isogeny α in our standard form (so $\operatorname{char}(k) \neq 2$).

Proof. If we put α in standard form $\left(\frac{u}{v}, \frac{s}{t}y\right)$ then the polynomial $v(x)$ has at most $\deg v$ distinct roots in $\bar{k}$, each of which can occur as the x-coordinate of at most two points on the elliptic curve E_1. □

Remark 4.30. Note that this corollary would not be true if we included the zero morphism in our definition of an isogeny.

One can also use the standard form of an isogeny $\alpha\colon E_1 \to E_2$ to show that α is surjective as a map from $E_1(\bar{k})$ to $E_2(\bar{k})$; see [7, Thm. 2.22].[3] But we already know that this applies to any non-constant morphism of curves (and even included surjectivity in our original definition of an isogeny), so we won't bother to prove this.

4.5 Degree and separability

We now define two important invariants of an isogeny that can be easily determined when it is in standard form.

Definition 4.31. Let $\alpha(x,y) = \left(\frac{u(x)}{v(x)}, \frac{s(x)}{t(x)}y\right)$ be an isogeny in standard form. The *degree* of α is $\deg\alpha := \max\{\deg u, \deg v\}$, and we say that α is *separable* if the derivative of $\frac{u(x)}{v(x)}$ is nonzero; otherwise we say that α is *inseparable*.

As noted earlier, the polynomials u, v, s, t are uniquely determined up to a scalar factor, so the degree and separability of α are intrinsic properties that do not depend on its representation as a rational map.

Remark 4.32. The degree and separability of an isogeny can be defined in a way that is more obviously intrinsic using function fields. If $\alpha\colon E_1 \to E_2$ is an isogeny of elliptic curves defined over k then it induces an injection of function fields

$$\alpha^*\colon k(E_2) \to k(E_1)$$

that sends f to $f \circ \alpha$ (notice the direction of this map; the categorical equivalence between smooth projective curves and their function fields is contravariant). The degree of α is then

[3] The theorem in [7] assumes that α is an endomorphism but the proof works for any isogeny.

the degree of $k(E_1)$ as an extension of the subfield $\alpha^*(k(E_2))$; this degree is finite because both are finite extensions of a purely transcendental extension of k. The isogeny α is then said to be separable if this field extension is separable (and is inseparable otherwise). This approach has the virtue of generality, but it is not as easy to apply explicitly. Our definition is equivalent, but we won't prove this.

Let us now return to the three examples that we saw earlier.

- The standard form of the negation map is $\alpha(x, y) = (x, -y)$. It is separable and has degree 1.
- The standard form of the multiplication-by-2 isogeny on $y^2 = x^3 + Ax + B$ is

$$\alpha(x,y) = \left(\frac{x^4 - 2Ax^2 - 8Bx + A^2}{4(x^3 + Ax + B)}, \frac{x^6 + 5Ax^4 + 20Bx^3 - 5A^2x^2 - 4ABx - A^3 - 8B^2}{8(x^3 + Ax + B)^2} y\right).$$

  It is separable and has degree 4.
- The standard form of the Frobenius endomorphism of $E\colon y^2 = f(x)$ over $\mathbb{F}_q$ is

$$\pi_E(x,y) = \left(x^q, f(x)^{(q-1)/2} y\right).$$

  We have used the curve equation to replace y^q with $f(x)^{(q-1)/2}y$; note that q is odd because we are not in characteristic 2. The Frobenius endomorphism is inseparable, because $(x^q)' = qx^{q-1} = 0$ in $\mathbb{F}_q$ (since q is a multiple of the characteristic p), and it has degree q.

4.6 Field extensions

Most of the material in this section can be found in any standard introductory algebra text, such as [1, 3]. We will occasionally need results in slightly greater generality than you may have seen before, and here we may reference [4, 5].

We start in the general setting of an arbitrary field extension L/k with no restrictions on k or L. The fields k and L necessarily have the same prime field (the subfield of k generated by the multiplicative identity), and therefore the same characteristic. The *degree* of the extension L/k, denoted $[L\colon k]$, is the dimension of L as a k-vector space; this is a cardinal number, which need not be finite. If we have a tower of fields $k \subseteq L \subseteq M$, then

$$[M : k] = [M : L][L : k],$$

where the RHS is a product of cardinals.[4] When $[L : k]$ is finite we say that L/k is a *finite extension*.

An element $\alpha \in L$ is said to be *algebraic* over k if it is the root of a polynomial in $k[x]$, and otherwise it is *transcendental* over k. The extension L/k is *algebraic* if every element of L is algebraic over k, and otherwise it is transcendental. If M/L and L/k are both algebraic extensions, so is M/k. A necessary and sufficient condition for L/k to be algebraic is that L be equal to the union of all finite extensions of k contained in L; in particular, every finite extension is algebraic.

[4] Recall that a cardinal number is an equivalence class of equipotent sets (sets that can be put in bijection). The product of $n_1 = \#S_1$ and $n_2 = \#S_2$ is $n_1 n_2 = \#(S_1 \times S_2)$ and the sum is the cardinality of the disjoint union: $n_1 + n_2 = \#(S_1 \sqcup S_2)$. But we shall be primarily interested in finite cardinals (natural numbers).

The subset of L consisting of the elements that are algebraic over k forms a field called the *algebraic closure* of k in L. A field k is *algebraically closed* if every non-constant polynomial in $k[x]$ has a root in k; equivalently, k has no non-trivial algebraic extensions. For every field k there exists an extension $\bar{k}/k$ with $\bar{k}$ algebraically closed; such a $\bar{k}$ is called an *algebraic closure* of k, and all such $\bar{k}$ are isomorphic (but this isomorphism is not unique in general). Any algebraic extension L/k can be embedded into any algebraic closure of k, since every algebraic closure of L is also an algebraic closure of k.

Remark 4.33. When working with algebraic extensions of k it is convenient to view them all as subfields of some fixed algebraic closure $\bar{k}$ (there is in general no canonical choice). The key point is that we can always (not necessarily uniquely) embed any algebraic extension of L/k in our chosen $\bar{k}$, and if we have another extension M/L, our embedding of L into $\bar{k}$ can always be extended to an embedding of M into $\bar{k}$.

A set $S \subseteq L$ is said to be *algebraically independent* (over k) if for every finite subset $\{s_1, \dots, s_n\}$ of S and every nonzero polynomial $f \in k[x_1, \dots, x_n]$ we have

$$f(s_1, \dots, s_n) \neq 0.$$

Note that this means the empty set is algebraically independent (just as the empty set is linearly independent in any vector space). An algebraically independent set $S \subseteq L$ for which $L/k(S)$ is algebraic is called a *transcendence basis* for the extension L/k.

Theorem 4.34. *Every transcendence basis for L/k has the same cardinality.*

Proof. We will only prove this in the case that L/k has a finite transcendence basis (which includes all extensions of interest to us); see [4, Theorem 7.9] for the general case. Let $S = \{s_1, \dots, s_m\}$ be a smallest transcendence basis and let $T = \{t_1, \dots, t_n\}$ be any other transcendence basis, with $n \geq m$. The set $\{t_1, s_1, \dots, s_m\}$ must then algebraically dependent, since $t_1 \in L$ is algebraic over $k(S)$, and since t_1 is transcendental over k, some s_i, say s_1, must be algebraic over $k(t_1, s_2, \dots, s_m)$. It follows that L is algebraic over $k(t_1, s_2, \dots, s_m)$, and the set $T_1 = \{t_1, s_2, \dots, s_m\}$ must be algebraically independent, otherwise it would contain a transcendence basis for L/k smaller than S. So T_1 is a transcendence basis for L/k of cardinality m that contains t_1.

Continuing in this fashion, for $i = 2, \dots, m$ we can iteratively construct transcendence bases T_i of cardinality m that contain $\{t_1, \dots, t_i\}$, until $T_m \subseteq T$ is a transcendence basis of cardinality m; but then we must have $T_m = T$, so $n = m$. □

Definition 4.35. The *transcendence degree* of a field extension L/K is the cardinality of any (hence every) transcendence basis for L/k.

Unlike extension degrees, which multiply in towers, transcendence degrees add in towers: for any fields $k \subseteq L \subseteq M$, the transcendence degree of M/k is the sum (as cardinals) of the transcendence degrees of M/L and L/k.

We say that the extension L/k is *purely transcendental* if $L = k(S)$ for some transcendence basis S for L/k. All purely transcendental extensions of k with the same transcendence degree are isomorphic. Every field extension L/k can be viewed as an algebraic extension of a purely transcendental extension: if S is a transcendence basis of L/k then $L/k(S)$ is an algebraic extension of the purely transcendental extension $k(S)/k$.

Remark 4.36. It is not the case that every field extension is a purely transcendental extension of an algebraic extension; indeed, most function fields are counterexamples.

The field extension L/k is said to be *simple* if $L = k(x)$ for some $x \in L$. A purely transcendental extension of transcendence degree 1 is obviously simple, but, less trivially, so is any finite separable extension (see below for the definition of separable); this is known as the primitive element theorem.

Remark 4.37. The notation $k(x)$ can be slightly confusing. If $x \in L$ is transcendental over k then $k(x)$ is isomorphic to the field of rational functions over k, in which case we may as well regard x as a variable. But if $x \in L$ is algebraic over k, then every rational expression $r(x)$ with nonzero denominator can be simplified to a polynomial in x of degree less than $n = [k(x) : k]$ by reducing modulo the minimal polynomial f of x (note that we can invert nonzero denominators modulo f); indeed, this follows from the fact that $\{1, x, \dots, x^{n-1}\}$ is a basis for the n-dimensional k-vector space $k(x)$.

4.6.1 Algebraic extensions

We now assume that L/k is algebraic and fix $\bar{k}$ so that $L \in \bar{k}$. The extension L/k is *normal* if it satisfies either of the equivalent conditions:

- every irreducible polynomial in $k[x]$ with a root in L splits completely in L;
- $\sigma(L) = L$ for all $\sigma \in \mathrm{Aut}(\bar{k}/k)$ (every automorphism of $\bar{k}$ that fixes k also fixes L).[5]

Even if L/k is not normal, there is always an algebraic extension M/L for which M/k is normal. The minimal such extension is called the *normal closure* of L/k; it exists because intersections of normal extensions are normal. It is not true in general that if L/k and M/L are normal extensions then so is M/k, but if $k \subseteq L \subseteq M$ is a tower of fields with M/k normal, then M/L is normal (but L/k need not be).

A polynomial $f \in k[x]$ is *separable* if any of the following equivalent conditions hold:

- the factors of f in $\bar{k}[x]$ are all distinct;
- f and f' have no common root in $\bar{k}$;
- $\gcd(f, f') = 1$ in $k[x]$.

An element $\alpha \in L$ is separable over k if any of the following equivalent conditions hold:

- α is a root of a separable polynomial $f \in k[x]$;
- the minimal polynomial of α is separable;
- $\mathrm{char}(k) = 0$ or $\mathrm{char}(k) = p > 0$ and the minimal polynomial of α is not of the form $g(x^p)$ for some $g \in k[x]$.

The elements of L that are separable over k form a field called the *separable closure* of k in L. The separable closure of k in its algebraic closure $\bar{k}$ is denoted k^{sep} and is simply called the separable closure of k. If $k \subseteq L \subseteq M$ then M/k is separable if and only if both M/L and L/k are separable.

Definition 4.38. A field k is *perfect* if any of the following equivalent conditions hold:

- $\mathrm{char}(k) = 0$ or $\mathrm{char}(k) = p > 0$ and $k = \{x^p : x \in k\}$ (k is fixed by Frobenius);
- every finite extension of k is separable over k;
- every algebraic extension of k is separable over k.

[5] Some authors write $\mathrm{Gal}(L/k)$ for $\mathrm{Aut}(L/k)$, others only use $\mathrm{Gal}(L/k)$ when L/k is known to be Galois; we will use the latter convention.

It is clear from the definition that finite fields and all fields of characteristic 0 are perfect, which includes most of the fields of interest to us in this course.

Example 4.39. The rational function field $k = \mathbb{F}_p(t)$ is not perfect. If we consider the finite extension $L = k(t^{1/p})$ obtained by adjoining a pth root of t to k, the minimal polynomial of $t^{1/p}$ is $x^p - t$, which is irreducible over k but not separable (its derivative is 0).

Definition 4.40. An algebraic extension L/k is *Galois* if it is both normal and separable, in which case we call $\mathrm{Gal}(L/k) = \mathrm{Aut}(L/k)$ the *Galois group* of L/k.

The extension k^{sep}/k is always normal: if an irreducible polynomial $f \in k[x]$ has a root α in k^{sep}, then (up to scalars) f is the minimal polynomial of α over k, hence separable over k, so all its roots lie in k^{sep}. Thus k^{sep}/k is a Galois extension and its Galois group

$$G_k \coloneqq \mathrm{Gal}(k^{\mathrm{sep}}/k)$$

is the *absolute Galois group* of k (we could also define G_k as $\mathrm{Aut}(\bar{k}/k)$, since the restriction map from $\mathrm{Aut}(\bar{k}/k)$ to $\mathrm{Gal}(k^{\mathrm{sep}}/k)$ is an isomorphism).

The *splitting field* of a polynomial $f \in k[x]$ is the extension of k obtained by adjoining all the roots of f (which lie in $\bar{k}$). Every splitting field is normal, and every finite normal extension of k is the splitting field of some polynomial over k; when k is a perfect field we can go further and say that L/k is a finite Galois extension if and only if it is the splitting field of some polynomial over k.

For finite Galois extensions M/k we always have $\#\mathrm{Gal}(M/k) = [M : k]$, and the fundamental theorem of Galois theory gives an inclusion-reversing bijection between subgroups $H \subseteq \mathrm{Gal}(M/k)$ and intermediate fields $k \subseteq L \subseteq M$ in which $L = M^H$ and $H = \mathrm{Gal}(M/L)$ (note that M/L is necessarily Galois). Beware that none of the statements in this paragraph necessarily apply to infinite Galois extensions; modifications are required.[6]

4.7 Algebraic sets

Let k be a perfect field and fix an algebraic closure $\bar{k}$.

Definition 4.41. The n-dimensional *affine space* $\mathbb{A}^n = \mathbb{A}^n_k$ over k is the set

$$\mathbb{A}^n \coloneqq \{(x_1, \dots, x_n) \in \bar{k}^n\},$$

equivalently, $\mathbb{A}^n$ is the vector space $\bar{k}^n$ regarded as a set. When k is clear from context we may just write $\mathbb{A}^n$. If $k \subseteq L \subseteq \bar{k}$, the set of *$L$-rational points* (or just L-points) in $\mathbb{A}^n$ is

$$\mathbb{A}^n(L) = \{(x_1, \dots, x_n) \in L^n\} = \mathbb{A}^n(\bar{k})^{G_L},$$

where $\mathbb{A}^n(\bar{k})^{G_L}$ denotes the set of points in $\mathbb{A}^n(\bar{k})$ fixed by $G_L \coloneqq \mathrm{Gal}(L^{\mathrm{sep}}/L)$. In particular, $\mathbb{A}^n(k) = \mathbb{A}^n(\bar{k})^{G_k}$.

Definition 4.42. If S is a set of polynomials in $\bar{k}[x_1, \dots, x_n]$, the set of points

$$Z_S \coloneqq \{P \in \mathbb{A}^n : f(P) = 0 \text{ for all } f \in S\},$$

is called an (affine) *algebraic set*. If $k \subseteq L \subseteq \bar{k}$, the set of L-rational points in Z_S is

$$Z_S(L) = Z_S \cap \mathbb{A}^n(L).$$

When S is a singleton $\{f\}$ we may write Z_f in place of $Z_{\{f\}}$.

[6] See Section 26.3 in the 18.785 Lecture notes for more details on infinite Galois extensions.

Note that if I is the $\bar{k}[x_1,\dots,x_n]$-ideal generated by S, then $Z_I = Z_S$, since $f(P) = g(P) = 0$ implies $(f+g)(P) = 0$ and $f(P) = 0$ implies $(fg)(P) = 0$. Thus we can always replace S by the ideal (S) that it generates, or by any set of generators for (S).

Example 4.43. We have $Z_\emptyset = Z_{(0)} = \mathbb{A}^n$ and $Z_{\{1\}} = Z_{(1)} = \emptyset$.

For any $S, T \subseteq A$ we have

$$S \subseteq T \quad \Longrightarrow \quad Z_T \subseteq Z_S,$$

but the converse need not hold, even if S and T are ideals: consider $T = (x_1)$ and $S = (x_1^2)$.

We now recall the notion of a noetherian ring and the Hilbert basis theorem.

Definition 4.44. A commutative ring R is *noetherian* if every R-ideal is finitely generated.[7] Equivalently, every infinite ascending chain of R-ideals

$$I_1 \subseteq I_2 \subseteq \cdots$$

eventually stabilizes, that is, $I_{n+1} = I_n$ for all sufficiently large n.

Theorem 4.45 (Hilbert basis theorem)**.** *If R is a noetherian ring, then so is $R[x]$.*

Proof. See [1, Theorem 14.6.7] or [3, Theorem 8.32]. □

Note that we can apply the Hilbert basis theorem repeatedly: if R is noetherian then so is $R[x_1]$, and so is $(R[x_1])[x_2] = R[x_1, x_2]$, ..., and so is $R[x_1,\dots,x_n]$. Like every field, $\bar{k}$ is a noetherian ring (it has just two ideals, so it certainly satisfies the ascending chain condition). Thus $A = \bar{k}[x_1,\dots,x_n]$ is noetherian, so every A-ideal is finitely generated. It follows that every algebraic set can be written in the form Z_S with S finite.

Definition 4.46. For an algebraic set $Z \subseteq \mathbb{A}^n$, the *ideal of* Z is the set

$$I(Z) = \{f \in \bar{k}[x_1,\dots,x_n] : f(P) = 0 \text{ for all } P \in Z\}.$$

The set $I(Z)$ is clearly an ideal, since it is closed under addition and under multiplication by elements of $\bar{k}[x_1,\dots,x_n]$, and we note that

$$Y \subseteq Z \implies I(Z) \subseteq I(Y)$$

and

$$I(Y \cup Z) = I(Y) \cap I(Z)$$

(both statements are immediate from the definition).

We have $Z = Z_{I(Z)}$ for every algebraic set Z, but it is not true that $I = I(Z_I)$ for every ideal I. As a counterexample, consider $I = (f^2)$ for some polynomial $f \in A$. In this case

$$I(Z_{(f^2)}) = (f) \neq (f^2).$$

In order to avoid this situation, we want to restrict our attention to *radical* ideals.

Definition 4.47. Let R be a commutative ring. For any R-ideal I we define

$$\sqrt{I} = \{x \in R : x^r \in I \text{ for some integer } r > 0\},$$

and say that I is a *radical ideal* if $I = \sqrt{I}$.

[7] The term "noetherian" refers to the German mathematician Emmy Noether.

Lemma 4.48. *For any ideal I in a commutative ring R, the set $\sqrt{I}$ is an ideal.*

Proof. Let $x \in \sqrt{I}$ with $x^r \in I$. For any $y \in R$ we have $y^r x^r = (xy)^r \in I$, so $xy \in \sqrt{I}$. If $y \in \sqrt{I}$ with $y^s \in I$, then every term in the sum

$$(x+y)^{r+s} = \sum_i \binom{r+s}{i} x^i y^{r+s-i}$$

is a multiple of either $x^r \in I$ or $y^s \in I$, hence lies in I, so $(x+y)^{r+s} \in I$ and $(x+y) \in \sqrt{I}$. □

Theorem 4.49 (Hilbert's *Nullstellensatz*)**.** *For every ideal $I \subseteq \bar{k}[x_1, \ldots, x_n]$ we have*

$$I(Z_I) = \sqrt{I}.$$

Proof. See [4, Theorem 7.1]. □

Nullstellensatz literally means "zero locus theorem". Theorem 4.49 is the strong form of the *Nullstellensatz*; it implies the *weak Nullstellensatz*.

Theorem 4.50 (*weak Nullstellensatz*)**.** *For any ideal $I \subsetneq \bar{k}[x_1, \ldots, x_n]$, the variety Z_I is nonempty.*

Proof. Suppose I is an ideal for which Z_I is the empty set. Then $I(Z_I) = (1)$, and by the strong *Nullstellensatz*, $\sqrt{I} = (1)$. But then $1^r = 1 \in I$, so $I = \bar{k}[x_1, \ldots, x_n]$. □

Note the importance of working over the algebraic closure $\bar{k}$. It is easy to find proper ideals I for which $Z_I(k) = \emptyset$ when k is not algebraically closed; consider $Z_{(x^2+y^2+1)}(\mathbb{Q})$ in $\mathbb{A}^2$. A useful corollary of the weak *Nullstellensatz* is the following.

Corollary 4.51. *The maximal ideals of the ring $\bar{k}[x_1, \ldots, x_n]$ are all of the form*

$$m_P = (x_1 - P_1, \ldots, x_n - P_n)$$

for some point $P = (P_1, \ldots, P_n)$ in $\mathbb{A}^n(\bar{k})$.

Proof. The evaluation map that sends $f \in \bar{k}[x_1, \ldots, x_n]$ to $f(P) \in \bar{k}$ is a surjective ring homomorphism with kernel m_P. Thus $\bar{k}[x_1, \ldots, x_n]/m_P \simeq \bar{k}$ is a field, hence m_P is a maximal ideal. If m is any maximal ideal in $\bar{k}[x_1, \ldots, x_n]$, then it is a proper ideal, and by the weak *Nullstellensatz* the algebraic set Z_m is nonempty and contains a point $P \in \mathbb{A}^n$. So $I(Z_m) \subseteq m_P$, but $m \subseteq I(Z_m) \subseteq m_P$ is maximal, so $m = m_P$. □

We also have the following corollary of Hilbert's *Nullstellensatz*.

Corollary 4.52. *There is a one-to-one inclusion-reversing correspondence between radical ideals $I \subseteq \bar{k}[x_1, \ldots, x_n]$ and algebraic sets $Z \subseteq \mathbb{A}^n(\bar{k})$ in which $I = I(Z)$ and $Z = Z_I$.*

Remark 4.53. It is hard to overstate the importance of Corollary 4.52; it is the basic fact that underlies nearly all of algebraic geometry. It tells us that the study of algebraic sets (geometric objects) is the same thing as the study of radical ideals (algebraic objects). It also suggests ways in which we might generalize our notion of an algebraic set: there is no reason to restrict ourselves to radical ideals in the ring $\bar{k}[x_1, \ldots, x_n]$, there are many other rings we might consider. This approach eventually leads to the more general notion of a *scheme*, which is the fundamental object in modern algebraic geometry.

Definition 4.54. An algebraic set is *irreducible* if it is nonempty and not the union of two smaller algebraic sets.

Theorem 4.55. *An algebraic set is irreducible if and only if its ideal is prime.*

Proof. ($\Rightarrow$) Let Y be an irreducible algebraic set and suppose $fg \in I(Y)$ for some $f, g \in A$. We will show that either $f \in I(Y)$ or $g \in I(Y)$ (and therefore $I(Y)$ is prime).

$$\begin{aligned} Y &\subseteq Z_{fg} = Z_f \cup Z_g \\ &= (Y \cap Z_f) \cup (Y \cap Z_g), \end{aligned}$$

and since Y is irreducible we must have either $Y = (Y \cap Z_f) = Z_f$ or $Y = (Y \cap Z_g) = Z_g$, hence either $f \in I(Y)$ or $g \in I(Y)$. Therefore $I(Y)$ is a prime ideal.

($\Leftarrow$) Now suppose $I(Y)$ is prime and that $Y = Y_1 \cup Y_2$. We will show that either $Y = Y_1$ or $Y = Y_2$. This will show that Y is irreducible, since Y must be nonempty ($I(Y) \neq A$ because $I(Y)$ is prime). We have

$$I(Y) = I(Y_1 \cup Y_2) = I(Y_1) \cap I(Y_2) \supseteq I(Y_1)I(Y_2),$$

and therefore $I(Y)$ divides/contains either $I(Y_1)$ or $I(Y_2)$, since $I(Y)$ is a prime ideal, but it is also contained in both $I(Y_1)$ and $I(Y_2)$, so either $I(Y) = I(Y_1)$ or $I(Y) = I(Y_2)$. Thus either $Y = Y_1$ or $Y = Y_2$, since algebraic sets with the same ideal must be equal. □

References

[1] Michael Artin, *Algebra*, 2nd edition, Pearson Education, 2011.

[2] Robin Hartshorne, *Algebraic geometry*, Graduate Texts in Mathematics **52**, Springer, 1977.

[3] Anthony W. Knapp, *Basic algebra*, Springer, 2006.

[4] Anthony W. Knapp, *Advanced algebra*, Springer, 2007.

[5] J. S. Milne, *Fields and Galois Theory*, 2012.

[6] J. H. Silverman, *The arithmetic of elliptic curves*, Graduate Texts in Mathematics **106**, second edition, Springer 2009.

[7] Lawrence C. Washington, *Elliptic curves: Number theory and cryptography*, second edition, Chapman and Hall/CRC, 2008.