Chapter 13

Fields

Our difficulty is not in the proofs, but in learning what to prove.

Emil Artin

1. EXAMPLES OF FIELDS

Much of the theory of fields has to do with a pair $F \subset K$ of fields, one contained in the other. In contrast with group theory, where subgroups play an important role, we usually consider K as an extension of F; that is, F is considered to be the basic field, and K is related to it. An *extension field* of F is a field which contains F as a subfield.

Here are the three most important classes of fields.

(1.1) **Number fields.** A number field K is a subfield of $\mathbb{C}$.

Any subfield of $\mathbb{C}$ contains 1, and hence it contains the field $\mathbb{Q}$ of rational numbers. So a number field is an extension of $\mathbb{Q}$. The number fields most commonly studied are *algebraic* number fields, all of whose elements are algebraic numbers (see Chapter 10, Section 1). We studied quadratic number fields in Chapter 11.

(1.2) **Finite fields.** A field having finitely many elements is called a finite field.

If K is a finite field, then the kernel of the unique homomorphism $\varphi: \mathbb{Z} \longrightarrow K$ is a prime ideal [Chapter 11 (7.15)], and since $\mathbb{Z}$ is infinite while K is finite, the kernel is not zero. Therefore it is generated by a prime integer p. The image of φ is isomorphic to the quotient $\mathbb{Z}/(p) = \mathbb{F}_p$. So K contains a subfield isomorphic to the prime field $\mathbb{F}_p$, and therefore it can be viewed as an extension of this prime field. We will describe all finite fields in Section 6.

(1.3) **Function fields.** Certain extensions of the field $F = \mathbb{C}(x)$ of rational functions are called function fields.

Function fields play an important role in the theory of analytic funtions and in algebraic geometry. Since we haven't seen them before, we will describe them briefly here. A function field can be defined by an irreducible polynomial in two variables, say $f(x,y) \in \mathbb{C}[x,y]$. The polynomial $f(x,y) = y^2 - x^3 + x$ is a good example. Given such a polynomial f, we may study the equation

$$f(x,y) = 0 \tag{1.4}$$

analytically, using it to define y "implicitly" as a function $y(x)$ of x as we learn to do in calculus. In our example, the function defined in this way is $y = \sqrt{x^3 - x}$. This function isn't single valued; it is determined only up to sign, but that isn't a serious difficulty. We won't have an explicit expression for such a function in general, but by definition, it satisfies the equation (1.4), that is,

$$f(x, y(x)) = 0. \tag{1.5}$$

On the other hand, the equation can also be studied algebraically. Let us interpret $f(x,y)$ as a polynomial in y whose coefficients are polynomials in x. Let F denote the field $\mathbb{C}(x)$ of rational functions in x. If f is not a polynomial in x alone, then since it is irreducible in $\mathbb{C}[x,y]$, it will be an irreducible element of $F[y]$ [Chapter 11 (3.9)]. Therefore the ideal generated by f in $F[y]$ is maximal [Chapter 11 (1.6)], and $F[y]/(f) = K$ is an extension field of F.

The analysis and the algebra are related, because both the implicitly defined function $y(x)$ and the residue $\overline{y}$ of y in $F[y]/(f)$ satisfy the equation $f(x,y) = 0$. In this way, the residue of y, and indeed all elements of K, can be interpreted as functions of the variable x. Because of this, such fields are called function fields. We will discuss function fields in Section 7.

2. ALGEBRAIC AND TRANSCENDENTAL ELEMENTS

Let K be an extension of a field F, and let α be an element of K. In analogy with the definition of algebraic numbers (Chapter 10, Section 1), α is said to be *algebraic over F* if it is the root of some nonzero polynomial with coefficients in F. Since the coefficients are from a field, we may assume that the polynomial is monic, say

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0, \quad \text{with } a_i \in F. \tag{2.1}$$

An element α is called *transcendental over F* if it is not algebraic over F, that is, if it is not a root of any such polynomial.

Note that the two properties, algebraic and transcendental, depend on the given field F. For example, the complex number $2\pi i$ is algebraic over the field of real numbers but transcendental over the field of rational numbers. Also, every element α of a field K is algebraic over K, because it is the root of the polynomial $x - \alpha$, which has coefficients in K.

The two possibilities for α can be described in terms of the substitution homomorphism

(2.2) $$\varphi\colon F[x] \longrightarrow K, \quad \text{which maps } f(x) \rightsquigarrow f(\alpha).$$

The element α is transcendental over F if φ is injective and algebraic over F otherwise, that is, if the kernel of φ is not zero.

Assume that α is algebraic over F. Since $F[x]$ is a principal ideal domain, ker φ is generated by a single element $f(x)$, the monic polynomial of lowest degree having α as a root. Since K is a field, we know that $f(x)$ must be an irreducible polynomial [Chapter 11 (7.15)], and in fact it will be the only irreducible monic polynomial in the ideal. Every other element of the ideal is a multiple of $f(x)$. We will call this polynomial f the *irreducible polynomial for α over F.*

It is important to note that this irreducible polynomial f depends on F as well as on α, because irreducibility of a polynomial depends on the field. For example, let $F = \mathbb{Q}[i]$, and let α be the complex number $\sqrt{i} = \frac{1}{2}\sqrt{2}(1 + i)$. The irreducible polynomial for α over $\mathbb{Q}$ is $x^4 + 1$, but this polynomial factors in the field F: $x^4 + 1 = (x^2 + i)(x^2 - i)$. The irreducible polynomial for α over F is $x^2 - i$. When there are several fields around, we must be careful to make it clear to which field we refer. To say that a polynomial is irreducible is ambiguous. It is better to say that f is *irreducible over F*, or that it is an *irreducible element of $F[x]$*.

The field extension of F which is generated by an element $\alpha \in K$ will be denoted by $F(\alpha)$:

(2.3) *$F(\alpha)$ is the smallest field containing F and α.*

More generally, if $\alpha_1,\dots,\alpha_n$ are elements of an extension field K of F, then the notation $F(\alpha_1,\dots,\alpha_n)$ will stand for the smallest subfield K which contains these elements.

As in Chapter 10, we denote the *ring* generated by α over F by $F[\alpha]$. It consists of all elements of K which can be written as polynomials in α with coefficients in F:

(2.4) $$a_n\alpha^n + \cdots + a_1\alpha + a_0, \quad a_i \in F.$$

The field $F(\alpha)$ is isomorphic to the field of fractions of $F[\alpha]$. Its elements are ratios of elements of the form (2.4) [see Chapter 10 (6.7)].

(2.5) **Proposition.** If α is transcendental over F, then the map $F[x] \longrightarrow F[\alpha]$ is an isomorphism, and hence $F(\alpha)$ is isomorphic to the field $F(x)$ of rational functions. □

This simple fact has the consequence that the field extensions $F(\alpha)$ are isomorphic for all transcendental elements α, because they are all isomorphic to the field of rational functions $F(x)$. For instance, π and e are both transcendental over $\mathbb{Q}$ (though we have not proved that they are). Therefore $\mathbb{Q}(\pi)$ and $\mathbb{Q}(e)$ are isomorphic fields, the isomorphism carrying π to e. This is rather surprising at first glance. The isomorphism is not continuous when the fields are regarded as subfields of the real numbers.

The situation is quite different if α is algebraic:

(2.6) **Proposition.**

(a) Suppose that α is algebraic over F, and let $f(x)$ be its irreducible polynomial over F. The map $F[x]/(f) \longrightarrow F[\alpha]$ is an isomorphism, and $F[\alpha]$ is a field. Thus $F[\alpha] = F(\alpha)$.

(b) More generally, let $\alpha_1, \dots, \alpha_n$ be algebraic elements of a field extension K of F. Then $F[\alpha_1, \dots, \alpha_n] = F(\alpha_1, \dots, \alpha_n)$.

Proof. Let φ be the map (2.2), with $K = F(\alpha)$. Since $f(x)$ generates ker φ, we know that $F[x]/(f)$ is isomorphic to the image of φ [Chapter 10 (3.1)], which is $F[\alpha]$. Since f is irreducible, it generates a maximal ideal [Chapter 11 (1.6)]. This shows that $F[\alpha]$ is a field. Since $F(\alpha)$ is isomorphic to the fraction field of $F[\alpha]$, it is equal to $F[\alpha]$. We leave the proof of the second part as an exercise. □

(2.7) **Proposition.** Let α be an algebraic element over F, and let $f(x)$ be its irreducible polynomial. Suppose $f(x)$ has degree n. Then $(1, \alpha, \dots, \alpha^{n-1})$ is a basis for $F[\alpha]$ as a vector space over F.

Proof. This proposition is a special case of (5.7) in Chapter 10. □

It may not be easy to tell whether or not two algebraic elements α, β generate isomorphic fields, though we can use Proposition (2.7) to give a *necessary* condition: Their irreducible polynomials over F must have the same degree, because this degree is the dimension of the field extension as an F-vector space. This is obviously not a sufficient condition. For example, all the imaginary quadratic fields studied in Chapter 11 are obtained by adjoining elements δ whose irreducible polynomials $x^2 - d$ have degree 2, but they aren't all isomorphic. On the other hand, if α is a root of $x^3 - x + 1$, then $\beta = \alpha^2$ is a root of $x^3 - 2x^2 + x - 1$. The two fields $\mathbb{Q}(\alpha)$ and $\mathbb{Q}(\beta)$ are actually equal, though if we were presented only with the two polynomials, it might take us some time to notice how they are related.

What we can describe easily are the circumstances under which there is an isomorphism

$$F(\alpha) \xrightarrow{\sim} F(\beta) \tag{2.8}$$

which fixes F and sends α to β. The following proposition is fundamental to our understanding of field extensions:

(2.9) **Proposition.** Let $\alpha \in K$ and $\beta \in L$ be algebraic elements of two extension fields of F. There is an isomorphism of fields

$$\sigma \colon F(\alpha) \xrightarrow{\sim} F(\beta),$$

which is the identity on the subfield F and which sends $\alpha \rightsquigarrow \beta$ if and only if the irreducible polynomials for α and β over F are equal.

Proof. Assume that $f(x)$ is the irreducible polynomial for α and for β over F. We apply Proposition (2.6), obtaining two isomorphisms

$$F[x]/(f) \xrightarrow{\varphi} F[\alpha] \quad \text{and} \quad F[x]/(f) \xrightarrow{\psi} F[\beta].$$

The composed map $\sigma = \psi\varphi^{-1}$ is the required isomorphism. Conversely, if there is an isomorphism σ sending α to β which is the identity on F, and if $f(x) \in F[x]$ is a polynomial such that $f(\alpha) = 0$, then $f(\beta) = 0$ too [see Proposition (2.11)]. Hence the two elements have the same irreducible polynomial. □

(2.10) **Definition.** Let K and K' be two extensions of the same field F. An isomorphism $\varphi: K \longrightarrow K'$ which restricts to the identity on the subfield F is called an *isomorphism of field extensions*, or an *F-isomorphism*. Two extensions K, K' of a field F are said to be *isomorphic field extensions* if there exists an F-isomorphism $\varphi: K \longrightarrow K'$.

(2.11) **Proposition.** Let $\varphi: K \longrightarrow K'$ be an isomorphism of field extensions of F, and let $f(x)$ be a polynomial with coefficients in F. Let α be a root of f in K, and let $\alpha' = \varphi(\alpha)$ be its image in K'. Then α' is also a root of f.

Proof. Say that $f(x) = a_nx^n + \cdots + a_1x + a_0$. Then $\varphi(a_i) = a_i$ and $\varphi(\alpha) = \alpha'$. Since φ is a homomorphism, we can expand as follows:

$$\begin{aligned} 0 = \varphi(0) = \varphi(f(\alpha)) &= \varphi(a_n\alpha^n + \cdots + a_1\alpha + a_0) \\ &= \varphi(a_n)\varphi(\alpha)^n + \cdots + \varphi(a_1)\varphi(\alpha) + \varphi(a_0) \\ &= a_n\alpha'^n + \cdots + a_1\alpha' + a_0. \end{aligned}$$

This shows that α' is a root of f. □

For example, the polynomial $x^3 - 2$ is irreducible over $\mathbb{Q}$. Let α denote the real cube root of 2, and let $\zeta = e^{2\pi i/3}$ be a complex cube root of 1. The three complex roots of $x^3 - 2$ are α, $\zeta\alpha$, and $\zeta^2\alpha$. Therefore there is an isomorphism

(2.12) $$\mathbb{Q}(\alpha) \xrightarrow{\sim} \mathbb{Q}(\zeta\alpha)$$

sending α to $\zeta\alpha$. In this case the elements of $\mathbb{Q}(\alpha)$ are all real numbers, but $\mathbb{Q}(\zeta\alpha)$ is not a subfield of $\mathbb{R}$. To understand the isomorphism (2.12), we must stop viewing these fields as subfields of $\mathbb{C}$ and look only at their internal algebraic structure.

3. THE DEGREE OF A FIELD EXTENSION

An extension K of a field F can always be regarded as an F-vector space. Addition is the addition law in K, and scalar multiplication of an element α of K by an element c of F is defined to be the product $c\alpha$ formed by multiplying these two elements in

K. The dimension of K as an F-vector space is called the *degree* of the field extension $F \subset K$. The degree is the simplest invariant of an extension, but though simple, it is important. It will be denoted by

(3.1) $[K : F] =$ *dimension of* K*, as an* F*-vector space.*

For example, $\mathbb{C}$ has the $\mathbb{R}$-basis $(1, i)$, so $[\mathbb{C} : \mathbb{R}] = 2$.

A field extension $F \subset K$ is called a *finite extension* if its degree $[K : F]$ is finite. Extensions of degree 2 are also called *quadratic* extensions, those of degree 3 are called *cubic* extensions, and so on. The degree of an extension $F \subset K$ is 1 if and only if $F = K$.

The term *degree* comes from the case that $K = F(\alpha)$ is generated by one algebraic element α. In that case, K has the basis $(1, \alpha, \ldots, \alpha^{n-1})$, where n is the degree of the irreducible polynomial for α over F [Proposition (2.7)]. Thus we find the first important property of the degree:

(3.2) **Proposition.** If α is algebraic over F, then $[F(\alpha) : F]$ is the degree of the irreducible polynomial for α over F. □

This degree is also called the *degree of* α *over* F. Note that an element α has degree 1 over F if and only if it is an element of F, and α has degree ∞ if and only if it is transcendental over F.

Extensions of degree 2 are easy to describe.

(3.3) **Proposition.** Assume that the field F does not have characteristic 2, that is, that $1 + 1 \neq 0$ in F. Then any extension $F \subset K$ of degree 2 can be obtained by adjoining a square root: $K = F(\delta)$, where $\delta^2 = D$ is an element of F. Conversely, if δ is an element of an extension of F, and if $\delta^2 \in F$ but $\delta \notin F$, then $F(\delta)$ is a quadratic extension.

Proof. We first show that every quadratic extension is obtained by adjoining a root of a quadratic polynomial $f(x) \in F[x]$. To do this, we choose any element α of K which is not in F. Then $(1, \alpha)$ is a linearly independent set over F. Since K has dimension 2 as a vector space over F, $(1, \alpha)$ is a basis for K over F, and $K = F[\alpha]$. It follows that α^2 is a linear combination of $(1, \alpha)$, say $\alpha^2 = -b\alpha - c$, with $b, c \in F$. Then α is a root of $f(x) = x^2 + bx + c$.

Since $2 \neq 0$ in F, we can use the quadratic formula $\alpha = \frac{1}{2}(-b + \sqrt{b^2-4c})$ to solve the equation $x^2 + bx + c = 0$. This is proved by direct calculation. There are two choices for the square root, one of which gives our chosen root α. Let δ denote that choice: $\delta = \sqrt{b^2-4c} = 2\alpha + b$. Then δ is in K, and it also generates K over F. Its square is the discriminant $b^2 - 4c$, which is in F.

The last assertion of the proposition is clear. □

The second important property of the degree is that it is multiplicative in towers of fields.

(3.4) **Theorem.** Let $F \subset K \subset L$ be fields. Then $[L : F] = [L : K][K : F]$.

Proof. Let $\mathbf{B} = (y_1,\dots,y_n)$ be a basis for L as a K-vector space, and let $\mathbf{C} = (x_1,\dots,x_m)$ be a basis for K as an F-vector space. So $[L : K] = n$ and $[K : F] = m$. We will show that the set of mn products $\mathbf{P} = (\dots,x_iy_j,\dots)$ is a basis of L as an F-vector space, and this will prove the proposition. The same reasoning will work if $\mathbf{B}$ or $\mathbf{C}$ is infinite.

Let α be an element of L. Since $\mathbf{B}$ is a basis for L over K, we can write $\alpha = \beta_1y_1 + \cdots + \beta_ny_n$, with $\beta_j \in K$, in a unique way. Since $\mathbf{C}$ is a basis for K over F, each β_i can be expressed uniquely, as $\beta_j = a_{1j}x_1 + \cdots + a_{mj}x_m$, with $a_{ij} \in F$. Thus $\alpha = \Sigma_{i,j}a_{ij}x_iy_j$. This shows that $\mathbf{P}$ spans L as an F-vector space. We know that β_j is uniquely determined by α, and since $\mathbf{B}$ is a basis for K over F, the elements a_{ij} are uniquely determined by β_j. So they are uniquely determined by α. This shows that $\mathbf{P}$ is linearly independent, and hence that it is a basis for L over F. □

One important case of a tower of field extensions is that K is a given extension of F and α is an element of K. Then the field $F(\alpha)$ generated by α is an intermediate field:

$$F \subset F(\alpha) \subset K. \tag{3.5}$$

(3.6) **Corollary.** Let K be an extension of F, of finite degree n. Let α be an element of K. Then α is algebraic over F, and its degree divides n.

To see this, we apply Theorem (3.4) to the fields $F \subset F(\alpha) \subset K$ and use the fact that the degree of α over F is $[F(\alpha) : F]$ if α is algebraic, while $[F(\alpha) : F] = \infty$ if α is transcendental. □

Here are some sample applications:

(3.7) **Corollary.** Let K be a field extension of F of prime degree p, and let α be an element of K which is not in F. Then α has degree p over F, and $K = F(\alpha)$.

For, $p = [K : F] = [K : F(\alpha)][F(\alpha) : F]$. One of the terms on the right side is 1. Since $\alpha \notin F$, it is not the second term, so $[K : F(\alpha)] = 1$ and $[F(\alpha) : F] = p$. Therefore $K = F(\alpha)$. □

(3.8) **Corollary.** Every irreducible polynomial in $\mathbb{R}[x]$ has degree 1 or 2.

We proved this in Chapter 11, Section 1, but let us derive it once more: Let g be an irreducible real polynomial. Then g has a root α in $\mathbb{C}$. Since $[\mathbb{C} : \mathbb{R}] = 2$, the degree of α over $\mathbb{R}$ divides 2, by (3.6). Therefore the degree of g is 1 or 2. □

(3.9) **Examples.**

(a) Let $\alpha = \sqrt[3]{2}$, $\beta = \sqrt[4]{5}$. Consider the field $L = \mathbb{Q}(\alpha,\beta)$ obtained by adjoining α and β to $\mathbb{Q}$. Then $[L : \mathbb{Q}] = 12$. For L contains the subfield $\mathbb{Q}(\alpha)$, which has degree 3 over $\mathbb{Q}$, because the irreducible polynomial for α over $\mathbb{Q}$ is $x^3 - 2$. Therefore 3 divides $[L : \mathbb{Q}]$. Similarly, L contains $\mathbb{Q}(\beta)$ and β has degree 4 over $\mathbb{Q}$, so 4 divides $[L : \mathbb{Q}]$. On the other hand, the degree of β over the field $\mathbb{Q}(\alpha)$ is at most 4, because β is a root of $x^4 - 5$, and this polynomial has coefficients in $\mathbb{Q}(\alpha)$. The chain of fields $L = \mathbb{Q}(\alpha, \beta) \supset \mathbb{Q}(\alpha) \supset \mathbb{Q}$ shows that $[L : \mathbb{Q}]$ is at most 12. So $[L : \mathbb{Q}] = 12$.

(b) It follows by reducing modulo 2 that the polynomial $f(x) = x^4 + 2x^3 + 6x^2 + x + 9$ is irreducible over $\mathbb{Q}$ [Chapter 11 (4.3)]. Let γ be a root of $f(x)$. Then there is no way to express $\alpha = \sqrt[3]{2}$ rationally in terms of γ, that is, $\alpha \notin \mathbb{Q}(\gamma)$. For $[\mathbb{Q}(\alpha) : \mathbb{Q}] = 3$, $[\mathbb{Q}(\gamma) : \mathbb{Q}] = 4$, and 3 does not divide 4. So we can't have $\mathbb{Q}(\gamma) > \mathbb{Q}(\alpha)$. On the other hand, since i has degree 2 over $\mathbb{Q}$, it is not so easy to decide whether i is in $\mathbb{Q}(\gamma)$. (In fact, it is not.) □

The next two theorems state the most important abstract consequences of the multiplicative property of degrees.

(3.10) **Theorem.** Let K be an extension of F. The elements of K which are algebraic over F form a subfield of K.

Proof. Let α, β be algebraic elements of K. We must show that $\alpha + \beta$, $\alpha\beta$, $-\alpha$, and α^{-1} (if $\alpha \neq 0$) are algebraic too. We note that since α is algebraic, $[F(\alpha) : F] < \infty$. Moreover, β is algebraic over F, and hence it is also algebraic over the bigger field $F(\alpha)$. Therefore the field $F(\alpha, \beta)$, which is generated over $F(\alpha)$ by β, is a finite extension of $F(\alpha)$, that is, $[F(\alpha, \beta) : F(\alpha)] < \infty$. By Theorem (3.4), $[F(\alpha, \beta) : F]$ is finite too. Therefore every element of $F(\alpha, \beta)$ is algebraic over F (3.6). The elements $\alpha + \beta$, $\alpha\beta$, etc. all lie in $F(\alpha, \beta)$, so they are algebraic. This proves that the algebraic elements form a field. □

Suppose for example that $\alpha = \sqrt{a}$, $\beta = \sqrt{b}$, where $a, b \in F$. Let us determine a polynomial having $\gamma = \alpha + \beta$ as a root. To do this, we compute the powers of γ, and we use the relations $\alpha^2 = a$, $\beta^2 = b$ to simplify when possible. Then we look for a linear relation among the powers:

$$\gamma^2 = \alpha^2 + 2\alpha\beta + \beta^2 = (a+b) + 2\alpha\beta$$

$$\gamma^4 = (a+b)^2 + 4(a+b)\alpha\beta + 4\alpha^2\beta^2 = (a^2+6ab+b^2) + 4(a+b)\alpha\beta.$$

We won't need the other powers because we can eliminate $\alpha\beta$ from these two equations to obtain the equation $\gamma^4 - 2(a+b)\gamma^2 + (a-b)^2 = 0$. Thus γ is a root of the polynomial

$$g(x) = x^4 - 2(a+b)x^2 + (a-b)^2,$$

which has coefficients in F, as required.

This method of undetermined coefficients will always produce a polynomial having an element such as $\alpha + \beta$ as a root, if the irreducible polynomials for α and β are known. Suppose that the degrees of two elements α, β are d_1, d_2, and let $n = d_1 d_2$. Any element of $F(\alpha, \beta)$ is a linear combination, with coefficients in F, of the n monomials $\alpha^i\beta^j$, $0 \leq i < d_1$, $0 \leq j < d_2$. This is because $F(\alpha, \beta) = F[\alpha, \beta]$ (2.6), and these monomials span $F[\alpha, \beta]$. Given an element $\gamma \in F(\alpha, \beta)$, we write the powers $1, \gamma, \gamma^2, \dots, \gamma^n$ as linear combinations of these monomials, with coefficients in F. Since there an $n + 1$ of the powers γ^ν and only n monomials $\alpha^i\beta^j$, the powers are linearly dependent. A linear dependence relation determines a polynomial with coefficients in F of which γ is a root.

But there is one point which complicates matters. Let $g(x)$ be the polynomial having γ as a root which we find in this way. This polynomial may be reducible. For instance, it may happen that γ is actually in the field F, though α, β aren't in F. If so, the method we described is unlikely to produce its irreducible equation $x - \gamma$. It is harder to determine the *irreducible* polynomial for γ over F. □

An extension K of a field F is called an *algebraic extension*, and K is said to be *algebraic over F*, if all its elements are algebraic.

(3.11) **Theorem.** Let $F \subset K \subset L$ be fields. If L is algebraic over K and K is algebraic over F, then L is algebraic over F.

Proof. We need to show that every element $\alpha \in L$ is algebraic over F. We are given that α is algebraic over K, hence that some equation of the form

$$\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0 = 0$$

holds, with $a_0, \dots, a_{n-1} \in K$. Therefore α is algebraic over the field $F(a_0, \dots, a_{n-1})$ generated by $a_0, \dots, a_{n-1}$ over F. Note that each coefficient a_i, being in K, is algebraic over F. We consider the chain of fields

$$F \subset F(a_0) \subset F(a_0, a_1) \subset \cdots \subset F(a_0, a_1, \dots, a_{n-1}) \subset F(a_0, a_1, \dots, a_{n-1}, \alpha)$$

obtained by adjoining the elements $a_0, \dots, a_{n-1}, \alpha$ in succession. For each i, a_{i+1} is algebraic over $F(a_0, \dots, a_i)$ because it is algebraic over F. Also, α is algebraic over $F(a_0, a_1, \dots, a_{n-1})$. So each extension in the chain is finite. By Theorem (3.4), the degree of $F(a_0, a_1, \dots, a_{n-1}, \alpha)$ over F is finite. Therefore by Corollary (3.6) α is algebraic over F. □

4. CONSTRUCTIONS WITH RULER AND COMPASS

There are famous theorems which assert that certain geometric constructions, such as trisection of an angle, can not be done with ruler and compass alone. We will now use the concept of degree of a field extension to prove some of them.

Here are the rules for basic ruler and compass construction:

(4.1)

(a) Two points in the plane are given to start with. These points are considered to be *constructed*.

(b) If two points have been constructed, we may draw the line through them, or draw a circle with center at one point and passing through the other. Such lines and circles are then considered to be *constructed*.

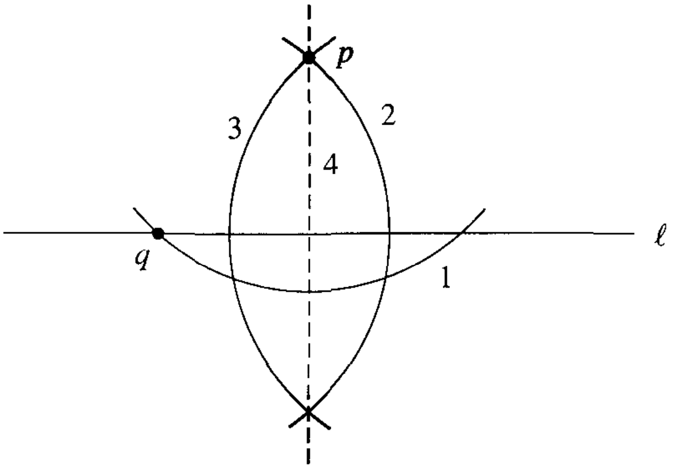
(c) The points of intersection of lines and circles which have been constructed are considered to be *constructed.*

Note that our ruler may be used only to draw straight lines through constructed points. We are not allowed to use it for measurement. Sometimes it is referred to as a "straight-edge" to make this point clear.

We will describe all possible constructions, beginning with some familiar ones. In each figure, the lines and circles are to be drawn in the order indicated.

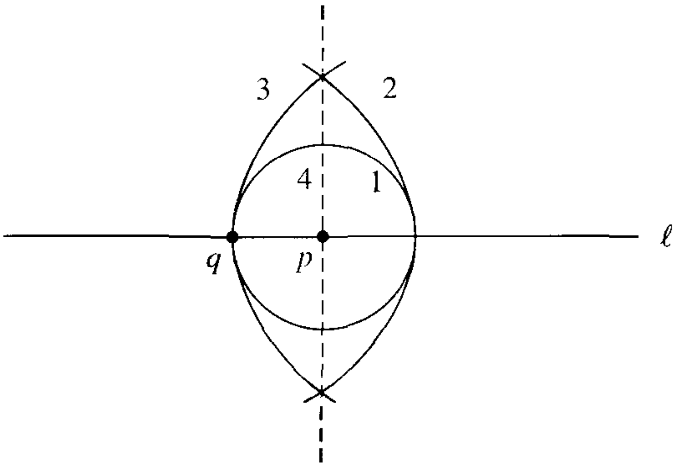
(4.2) **Construction.** Draw a line through a constructed point p and perpendicular to a constructed line ℓ.

Case 1: $p \notin \ell$

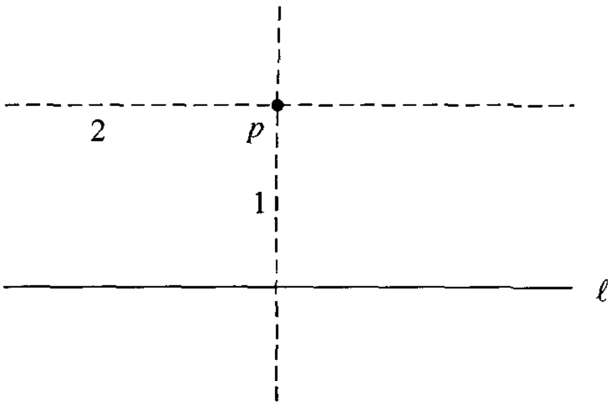


This construction works with any point $q \in \ell$ which is not on the perpendicular. However, we had better not choose points arbitrarily, because if we do we'll have difficulty keeping track of which points we have constructed and which ones are merely artifacts of an arbitrary choice. Whenever we want an arbitrary point, we will construct a particular one for the purpose.

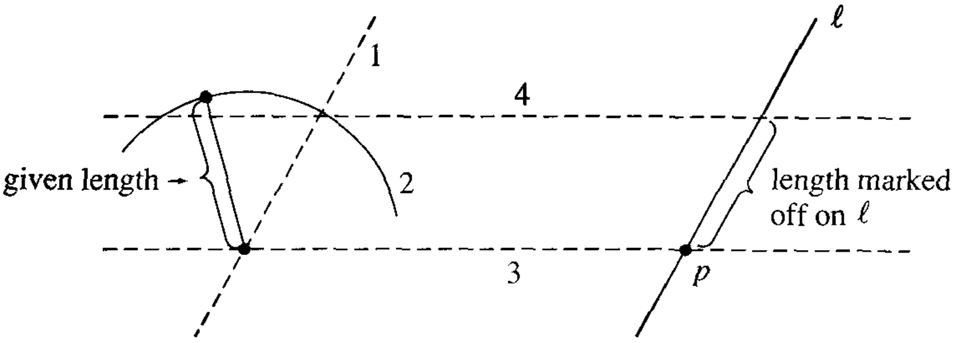
Case 2: $p \in \ell$



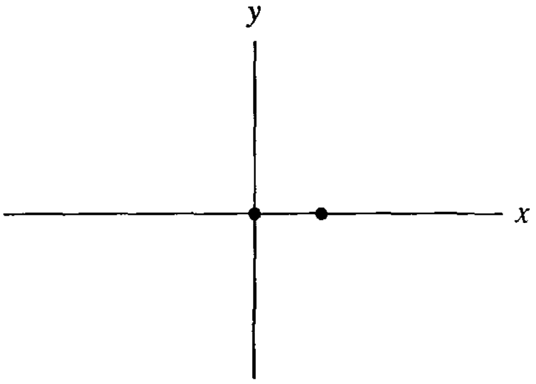
(4.3) **Construction.** Draw a line parallel to ℓ and passing through p. Apply Cases 1 and 2 above:



(4.4) **Construction.** Mark off a length defined by two points onto a constructed line ℓ, starting at a constructed point $p \in \ell$. Use construction of parallels.



These constructions allow us to introduce Cartesian coordinates into the plane so that the two points which are given to us to start have coordinates $(0, 0)$ and $(0, 1)$. Other choices of coordinate systems could be used, but they lead to equivalent theories.



We will call a real number a *constructible* if its absolute value $|a|$ is the distance between two constructible points, the unit length being the distance between the points given originally.

(4.5) **Proposition.** A point $p = (a, b)$ is constructible if and only if its Cartesian coordinates a and b are constructible numbers.

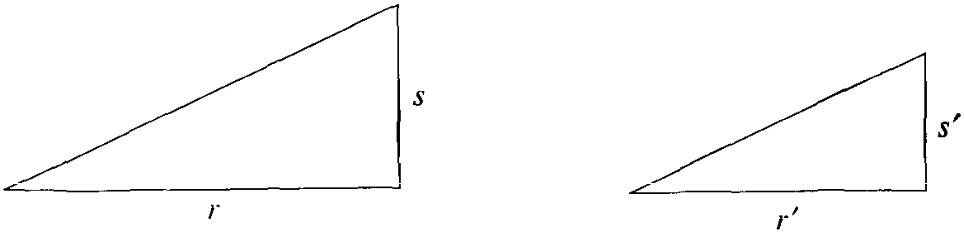
Proof. This follows from the above constructions. Given a point p, we can construct its coordinates by dropping perpendiculars to the axes. Conversely, if a and b are given constructible numbers, then we can construct the point p by marking a, b off on the two axes using (4.4) and erecting perpendiculars. □

(4.6) **Proposition.** The constructible numbers form a subfield of $\mathbb{R}$.

Proof. We will show that if a and b are positive constructible numbers, then $a + b$, ab, $a - b$, (if $a > b$), and a^{-1} (if $a \neq 0$) are also constructible. The closure in case a or b is negative follows easily.

Addition and subtraction are done by marking lengths on a line, using Construction (4.4).

For multiplication, we use similar right triangles:



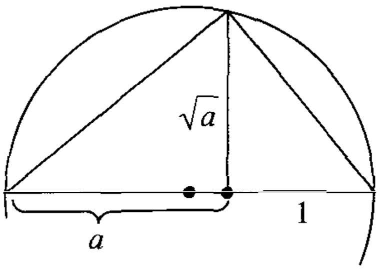
Given one triangle and one side of a second triangle, the second triangle can be constructed by parallels.

To construct the product ab, we take $r = 1$, $s = a$, and $r' = b$. Then since $r/s = r'/s'$, it follows that $s' = ab$. To construct a^{-1}, we take $r = a$, $s = 1$, and $r' = 1$. Then $s' = a^{-1}$. □

(4.7) **Proposition.** If a is a positive constructible number, then so is $\sqrt{a}$.

Proof. We use similar triangles again. We must construct them so that $r = a$, $r' = s$, and $s' = 1$. Then $s = r' = \sqrt{a}$.

How to make the construction is less obvious this time, but we can use inscribed triangles in a circle. A triangle inscribed into a circle, with a diameter as its hypotenuse, is a right triangle. This is a theorem of high school geometry. It can be checked using the equation for a circle and Pythagoras's theorem. So we draw a circle whose diameter is $1 + a$ and proceed as in the figure below. Note that the large triangle is divided into two similar triangles.



(4.8) **Proposition.** Suppose four points are given, whose coordinates are in a subfield F of $\mathbb{R}$. Let A, B be lines or circles drawn using the given points. Then the points of intersection of A and B have coordinates of F, or in a field of the form $F(\sqrt{r})$, where r is a positive number in F.

Proof. The line through (a_0, b_0), (a_1, b_1) has the linear equation

$$(a_1 - a_0)(y - b_0) = (b_1 - b_0)(x - a_0).$$

The circle with center (a_0, b_0) and passing through (a_1, b_1) has the quadratic equation

$$(x - a_0)^2 + (y - b_0)^2 = (a_1 - a_0)^2 + (b_1 - b_0)^2.$$

The intersection of two lines can be found by solving two linear equations whose coefficients are in F. So its coordinates are in F too. To find the intersection of a line and a circle, we use the equation of the line to eliminate one variable from the equation of the circle, obtaining a quadratic equation in one unknown. This quadratic equation has solutions in the field $F(\sqrt{D})$, where D is the discriminant, which is an element of F. If $D < 0$, the line and circle do not intersect.

Consider the intersection of two circles, say

$$(x - a_1)^2 + (y - b_1)^2 = r_1^2 \quad \text{and} \quad (x - a_2)^2 + (y - b_2)^2 = r_2^2,$$

where $a_i, b_i, r_i \in F$. In general, the solution of a pair of quadratic equations in two variables requires solving an equation of degree 4. In this case we are lucky: The difference of the two quadratic equations is a linear equation which we can use to eliminate one variable, as before. □

(4.9) **Theorem.** Let $a_1, \dots, a_m$ be constructible real numbers. There is a chain of subfields $\mathbb{Q} = F_0 \subset F_1 \subset F_2 \subset \cdots \subset F_n = K$ such that

(i) K is a subfield of $\mathbb{R}$;
(ii) $a_1, \dots, a_m \in K$;
(iii) for each $i = 0, \dots, n - 1$, the field F_{i+1} is obtained from F_i by adjoining the square root of a positive number $r_i \in F_i$, which is not a square in F_i.

Conversely, let $\mathbb{Q} = F_0 \subset F_1 \subset \cdots \subset F_n = K$ be a chain of subfields of $\mathbb{R}$ which satisfies (iii) . Then every element of K is constructible.

Proof. We introduced coordinates so that the points originally given have coordinates in $\mathbb{Q}$. The process of constructing the numbers a_i involves drawing lines and circles and taking their intersections. So the first assertion follows by induction from Proposition (4.8). Conversely, if such a tower of fields is given, then its elements are constructible, by Propositions (4.6) and (4.7). □

(4.10) **Corollary.** If a is a constructible real number, then it is algebraic, and its degree over $\mathbb{Q}$ is a power of 2.

For, in the chain of fields (4.9), the degree of F_{i+1} over F_i is 2, and hence $[K : \mathbb{Q}] = 2^n$. Corollary (3.6) tells us that the degree of a divides 2^n, hence that it is a power of 2. □

The converse of Corollary (4.10) is false. There exist real numbers a which have degree 4 over $\mathbb{Q}$ but which are not constructible. We will be able to prove this later, using Galois theory.

We can now prove the impossibility of certain geometric constructions. Our method will be to show that if a certain construction were possible, then it would also be possible to construct an algebraic number whose degree over $\mathbb{Q}$ is not a power of 2. This would contradict (4.10).

Let us discuss trisection of the angle as the first example. We must pose the problem carefully, because many angles, 45° for instance, can be trisected. The customary way to state the problem is to ask for a single method of construction which will work for *any given angle*.

To be as specific as possible, let us say that an angle θ is *constructible* if its cosine $\cos\theta$ is constructible. Other equivalent definitions are possible. For example, with this definition, θ is constructible if and only if the line which passes through the origin and meets the x-axis in the angle θ is constructible. Or, θ is constructible if and only if it is possible to construct any two lines meeting in an angle θ.

Now just giving an angle θ (say by marking off its cosine on the x-axis) provides us with new information which may be used in a hypothetical trisection. To analyze the consequences of this new information, we should start over and determine all constructions which can be made when, in addition to two points, one more length ($= \cos\theta$) is given at the start. We would rather not take the time to do this, and there is a way out. We will exhibit a particular angle θ with these properties:

(4.11) (i) θ is constructible, and

(ii) $\frac{1}{3}\theta$ is not constructible.

The first condition tells us that being given the angle θ provides no new information for us: If the angle θ can be trisected when given, it can also be trisected without being given. The second condition tells us that there is no general method of trisection, because there is no way to trisect θ.

The angle $\theta = 60°$ does the job. A 60° angle is constructible because $\cos 60° = \frac{1}{2}$. On the other hand, it is impossible to construct a 20° angle. To show this, we will show that $\cos 20°$ is an algebraic number of degree 3 over $\mathbb{Q}$. Then Corollary (4.10) will show that $\cos 20°$ is not constructible, hence that 60° can not be trisected.

The addition formulas for sine and cosine can be used to prove the identity

$$\cos 3\theta = 4\cos^3\theta - 3\cos\theta. \tag{4.12}$$

Setting $\theta = 20°$ and $\alpha = \cos 20°$, we obtain the equation $\frac{1}{2} = 4\alpha^3 - 3\alpha$, or $8\alpha^3 - 6\alpha - 1 = 0$.

(4.13) **Lemma.** The polynomial $f(x) = 8x^3 - 6x - 1$ is irreducible over $\mathbb{Q}$.

Proof. It is enough to check for linear factors $ax + b$, where a, b are integers such that a divides 8, and $b = \pm 1$. Another way to prove irreducibility is to check that f has no root modulo 5. □

This lemma tells us that α has degree 3 over $\mathbb{Q}$, hence that it can not be constructed.

As another example, let us show that the regular 7-gon can not be constructed. This is similar to the above problem: The construction of 20° is equivalent to the construction of the 18-gon. Let θ denote the angle $2\pi/7$ and let $\zeta = \cos\theta + i\sin\theta$. Then ζ is a root of the equation $x^6 + x^5 + \cdots + 1 = 0$, which is irreducible [Chapter 11 (4.6)]. Hence ζ has degree 6 over $\mathbb{Q}$. If the 7-gon were constructible, then $\cos\theta$ and $\sin\theta$ would be constructible numbers, and hence they would lie in a real field extension of degree 2^n over $\mathbb{Q}$, by Theorem (4.9). Call this field K, and consider the extension $K(i)$. This extension has degree 2. Therefore $[K(i) : \mathbb{Q}] = 2^{n+1}$. But $\zeta = \cos\theta + i\sin\theta \in K(i)$. This contradicts the fact that the degree of ζ is 6 (3.6).

Notice that this argument is not special to the number 7. It applies to any prime integer p, provided only that $p - 1$, the degree of the irreducible polynomial $x^{p-1} + \cdots + x + 1$, is not a power of 2.

(4.14) **Corollary.** Let p be a prime integer. If the regular p-gon can be constructed by ruler and compass, then $p = 2^r + 1$ for some integer r. □

Gauss proved the converse: If a prime has the form $2^r + 1$, then the regular p-gon can be constructed. The regular 17-gon, for example, can be constructed with ruler and compass. We will learn how to prove this in the next chapter.

5. SYMBOLIC ADJUNCTION OF ROOTS

Up to this point, we have used subfields of the complex numbers as our examples. Abstract constructions are not needed to create these fields (except that the construction of $\mathbb{C}$ from $\mathbb{R}$ is abstract). We simply adjoin complex numbers to the rational numbers as desired and work with the subfield they generate. But finite fields and function fields are not subfields of a familiar, all-encompassing field analogous to $\mathbb{C}$, so these fields must be constructed. The fundamental tool for their construction is the adjunction of elements to a ring, which we studied in Section 5 of Chapter 10. It is applied here to the case that the ring we start with is a field F.

Let us review this construction. Given a polynomial $f(x)$ with coefficients in F, we may adjoin an element α satisfying the polynomial equation $f(\alpha) = 0$ to F. The abstract procedure is to form the polynomial ring $F[x]$ and then take the quotient ring

$$R' = F[x]/(f). \tag{5.1}$$

This construction always yields a ring R' and a homomorphism $F \longrightarrow R'$, such that the residue $\overline{x}$ of x satisfies the relation $f(\overline{x}) = 0$.

However, we want to construct not only a ring, but a field, and here the theory of polynomials over a field comes into play. Namely, that theory tells us that the principal ideal (f) is a maximal ideal if and only if f is irreducible [Chapter 11 (1.6)]. Therefore the ring R' will be a field if and only if f is an irreducible polynomial.

(5.2) **Lemma.** Let F be a field, and let f be an irreducible polynomial in $F[x]$. Then the ring $K = F[x]/(f)$ is an extension field of F, and the residue $\overline{x}$ of x is a root of $f(x)$ in K.

Proof. The ring K is a field because (f) is a maximal ideal. Also, the homomorphism $F \longrightarrow K$, which sends the elements of F to the residues of the constant polynomials, is injective, because F is a field. So we may identify F with its image, a subfield of K. The field K becomes an extension of F by means of this identification. Finally, $\overline{x}$ satisfies the equation $f(\overline{x}) = 0$, which means that it is a root of f. □

(5.3) **Proposition.** Let F be a field, and let $f(x)$ be a monic polynomial in $F[x]$ of positive degree. There exists a field extension K of F such that $f(x)$ factors into linear factors over K.

Proof. We use induction on the degree of f. The first case is that f has a root α in F, so that $f(x) = (x - \alpha)g(x)$ for some polynomial g. If so, we replace f by g, and we are done by induction. Otherwise, we choose an irreducible factor $g(x)$ of $f(x)$. By Lemma (5.2), there is a field extension of F, call it F_1, in which $g(x)$ has a root α. We replace F by F_1 and are thereby reduced to the first case. □

As we have seen, the polynomial ring $F[x]$ is an important tool for studying extensions of a field F. When we are working with two fields at the same time, there is an interplay between their polynomial rings. This interplay doesn't present serious difficulties, but instead of scattering the points which need to be mentioned about in the text, we have collected them here.

Notice that if K is an extension field of F, then the polynomial ring $K[x]$ contains $F[x]$ as subring. So computations which are made in the ring $F[x]$ are also valid in $K[x]$.

(5.4) **Proposition.** Let f and g be polynomials with coefficients in a field F, and let K be an extension field of F.

(a) Division with remainder of g by f gives the same answer, whether carried out in $F[x]$ or in $K[x]$.

(b) f divides g in $K[x]$ if and only if f divides g in $F[x]$.

(c) The monic greatest common divisor d of f and g is the same, whether computed in $F[x]$ or in $K[x]$.

(d) If f and g have a common root in K, then they are not relatively prime in $F[x]$. Conversely, if f and g are not relatively prime in $F[x]$, then there exists an extension field L in which they have a common root.

(e) If f is irreducible in $F[x]$ and if f and g have a common root in K, then f divides g in $F[x]$.

Proof. (a) Carry out the division in $F[x] : g = fq + r$. This equation also holds in the bigger ring $K[x]$, and further division of the remainder by f is not possible, because r has lower degree than f, or else it is zero.

(b) This is the case that the remainder is zero in (a).

(c) Let d, d' denote the monic greatest common divisors of f and g in $F[x]$ and in $K[x]$. Then d is also a common divisor in $K[x]$. So d divides d' in $K[x]$, by definition of d'. In addition, we know that d has the form $d = pf + qg$, for some elements $p, q \in F[x]$. Since d' divides f and g, it divides $pf + qg = d$ too. Thus d and d' are associates in $K[x]$, and, being monic, they are equal.

(d) Let α be a common root of f and g in K. Then $x - \alpha$ is a common divisor of f and g in $K[x]$. So their greatest common divisor in $K[x]$ is not 1. By (c), it is not 1 in $F[x]$ either. Conversely, if f and g have a common divisor d of degree > 0, then by (5.3), d has a root in some extension field L. This root will be a common root of f and g.

(e) If f is irreducible, then its only divisors in $F[x]$ are $1, f$, and their associates. Part (d) tells us that the greatest common divisor of f and g in $F[x]$ is not 1. Therefore it is f. □

The final topic of this section concerns the derivative $f'(x)$ of a polynomial $f(x)$. In algebra, the derivative is computed using the rules from calculus for differentiating polynomial functions. In other words, we define the derivative of x^n to be the polynomial nx^{n-1}, and if $f(x) = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$, then

$$f'(x) = na_nx^{n-1} + (n-1)a_{n-1}x^{n-2} + \cdots + a_1. \tag{5.5}$$

The integer coefficients in this formula are to be interpreted as elements of F by means of the homomorphism $\mathbb{Z} \longrightarrow F$ [Chapter 10 (3.18)]. So the derivative is a polynomial with coefficients in the same field. It can be shown that rules such as the product rule for differentiation hold.

Though differentiation is an algebraic procedure, there is no a priori reason to suppose that it has much algebraic significance; however, it does. For us, the most important property of the derivative is that it can be used to recognize multiple roots of a polynomial.

(5.6) **Lemma.** Let F be a field, let $f(x) \in F[x]$ be a polynomial, and let $\alpha \in F$ be a root of $f(x)$. Then α is a *multiple root,* meaning that $(x - \alpha)^2$ divides $f(x)$, if and only if it is a root of both $f(x)$ and $f'(x)$.

Proof. If α is a root of f, then $x - \alpha$ divides f: $f(x) = (x - \alpha)g(x)$. Then α is a root of g if and only if it is a multiple root of f. By the product rule for differentiation,

$$f'(x) = (x - \alpha)g'(x) + g(x).$$

Substituting $x = \alpha$ shows that $f'(\alpha) = 0$ if and only if $g(\alpha) = 0$. □

(5.7) **Proposition.** Let $f(x) \in F[x]$ be a polynomial. There exists a field extension K of F in which f has a multiple root if and only if f and f' are not relatively prime.

Proof. If f has a multiple root in K, then f and f' have a common root in K by Lemma (5.6), and so they are not relatively prime in K. Hence they are not relatively prime in F either. Conversely, if f and f' are not relatively prime, then they have a common root in some field extension K, hence f has a multiple root there. □

Here is one of the most important applications of the derivative to field theory:

(5.8) **Proposition.** Let f be an irreducible polynomial in $F[x]$. Then f has no multiple root in any field extension of F unless the derivative f' is the zero polynomial. In particular, if F is a field of characteristic zero, then f has no multiple root.

Proof. By the previous proposition, we must show that f and f' are relatively prime unless f' is the zero polynomial. Since f is irreducible, the only way that it can have a nonconstant factor in common with another polynomial g is for f to divide g (5.4e). And if f divides g, then $\deg g \geq \deg f$, or else $g = 0$. Now the degree of the derivative f' is less than the degree of f. So f and f' have no nonconstant factor in common unless $f' = 0$, as required. In a field of characteristic zero, the derivative of a nonconstant polynomial is not zero. □

The derivative of a nonconstant polynomial $f(x)$ may be identically zero if F has prime characteristic p. This happens when the exponent of every monomial occurring in f is divisible by p. A typical polynomial whose derivative is zero in characteristic 5 is

$$f(x) = x^{15} + ax^{10} + bx^5 + c,$$

where a, b, c can be arbitrary elements of F. Since the derivative of this polynomial is identically zero, its roots in any extension field are all multiple roots. Whether or not this polynomial is irreducible depends on F and on a, b, c.

6. *FINITE FIELDS*

In this section, we describe all fields having finitely many elements. We remarked in Section 1 that a finite field K contains one of the prime fields $\mathbb{F}_p$, and of course since K is finite, it will be finite-dimensional when considered as a vector space over this field. Let us denote $\mathbb{F}_p$ by F, and let r denote the degree $[K : F]$. As an F-vector space, K is isomorphic to the space F^r, and this space contains p^r elements. So the order of a finite field is always a power of a prime. It is customary to use the letter q for this number:

$$(6.1) \qquad q = p^r = |K|.$$

When referring to finite fields, p will always denote a prime integer and q a power of p, the number of elements, or *order*, of the field.

Fields with q elements are often denoted by $\mathbb{F}_q$. We are going to show that all fields with the same number of elements are isomorphic, so this notation is not too ambiguous. However, the isomorphism will not be unique when $r > 1$.

The simplest example of a finite field other than the prime field $\mathbb{F}_p$ is the field $K = \mathbb{F}_4$ of order 4. There is a unique irreducible polynomial $f(x)$ of degree 2 in $\mathbb{F}_2[x]$, namely

$$(6.2) \qquad f(x) = x^2 + x + 1$$

[see Chapter 11 (4.3)], and the field K is obtained by adjoining a root α of $f(x)$ to $F = \mathbb{F}_2$:

$$K \approx F[x]/(x^2 + x + 1).$$

The order of this field is 4 because α has degree 2, which tells us that K has dimension 2 as a vector space over the field F.

The set $(1, \alpha)$ forms a basis of K over F, so the elements of K are the four linear combinations of these two elements, with mod-2 coefficients 0, 1. They are

$$(6.3) \qquad \{0, 1, \alpha, 1 + \alpha\} = \mathbb{F}_4.$$

The element $1 + \alpha$ is the second root of the polynomial $f(x)$ in K. Computation in K is made using the relations $1 + 1 = 0$ and $\alpha^2 + \alpha + 1 = 0$. *Do not confuse the field* $\mathbb{F}_4$ *with the ring* $\mathbb{Z}/(4)$!

Here are the main facts about finite fields:

(6.4) **Theorem.** Let p be a prime, and let $q = p^r$ be a power of p, with $r \geq 1$.

(a) There exists a field of order q.

(b) Any two fields of order q are isomorphic.

(c) Let K be a field of order q. The multiplicative group $K^\times$ of nonzero elements of K is a cyclic group of order $q - 1$.

(d) The elements of K are roots of the polynomial $x^q - x$. This polynomial has distinct roots, and it factors into linear factors in K.

(e) Every irreducible polynomial of degree r in $\mathbb{F}_p[x]$ is a factor of $x^q - x$. The irreducible factors of $x^q - x$ in $\mathbb{F}_p[x]$ are precisely the irreducible polynomials in $\mathbb{F}_p[x]$ whose degree divides r.

(f) A field K of order q contains a subfield of order $q' = p^k$ if and only if k divides r.

The proof of this theorem is not very difficult, but since there are several parts, it will take some time. To motivate it, we will look at a few consequences first.

The striking aspect of (c) is that all nonzero elements of K can be listed as powers of a single suitably chosen one. This is not obvious, even for the prime field $\mathbb{F}_p$. For example, the residue of 3 is a generator of $\mathbb{F}_7^\times$. Its powers $3^0, 3^1 3^2, \ldots$ list the nonzero elements of $\mathbb{F}_7$ in the following order:

$$(6.5) \qquad \mathbb{F}_7^\times = \{1, 3, 2, 6, 4, 5\}.$$

As another example, 2 is a generator of $\mathbb{F}_{11}^{\times}$, and its powers list that group in the order

$$(6.6) \qquad \mathbb{F}_{11}^{\times} = \{1,2,4,8,5,10,9,7,3,6\}.$$

A generator for the cyclic group $\mathbb{F}_p^{\times}$ is called a *primitive element modulo p*. Note that the theorem does not tell us how to find a primitive element, only that one exists. Which residues modulo p are primitive elements is not well understood, but given a small prime p, we can find one by trial and error.

We now have two ways of listing the nonzero elements of $\mathbb{F}_p$, additively and multiplicatively:

$$(6.7) \qquad \mathbb{F}_p^{\times} = \{1,2,3,\ldots,p-1\} = \{1,\nu,\nu^2,\ldots,\nu^{p-2}\},$$

where ν is a primitive element modulo p. Depending on the context, one or the other list may be the best for computation.

Of course, the additive group $\mathbb{F}_p^{+}$ of the prime field is always a cyclic group of order p. Both the additive and multiplicative structures of the prime field are very simple: They are cyclic. But the field structure of $\mathbb{F}_p$, governed by the distributive law, fits the two together in a subtle way.

Part (e) of the theorem is also striking. It is the basis for many methods of factoring polynomials modulo p. Let us look at a few cases in which q is a power of 2 as examples:

(6.8) **Examples.**

(a) The elements of the field $\mathbb{F}_4$ are the roots of the polynomial

$$(6.9) \qquad x^4 - x = x(x-1)(x^2+x+1).$$

In this case, the irreducible factors of $x^4 - x$ in $\mathbb{Z}[x]$ happen to remain irreducible in $\mathbb{F}_2[x]$. Note that the factors of $x^2 - x$ appear here, because $\mathbb{F}_4$ contains $\mathbb{F}_2$.

Since we are working in characteristic 2, the signs are irrelevant: $x - 1 = x + 1$.

(b) The field $\mathbb{F}_8$ of order 8 has degree 3 over the prime field $\mathbb{F}_2$. Its elements are the eight roots of the polynomial

$$(6.10) \qquad x^8 - x = x(x-1)(x^3+x+1)(x^3+x^2+1), \quad \text{in } \mathbb{F}_2[x].$$

So the six elements in $\mathbb{F}_8$ which aren't in $\mathbb{F}_2$ fall into two classes: the three roots of $x^3 + x + 1$ and the three roots of $x^3 + x^2 + 1$.

The cubic factors of (6.10) are the two irreducible cubic polynomials of degree 3 in $\mathbb{F}_2[x]$ [see Chapter 11 (4.3)]. Notice that the irreducible factorization of this polynomial in the ring of integers is

$$(6.11) \qquad x^8 - x = x(x-1)(x^6+x^5+\cdots+x+1), \quad \text{in } \mathbb{Z}[x].$$

The third factor is reducible modulo 2.

To compute in the field $\mathbb{F}_8$, choose a root β of one of the cubics, say of $x^3 + x + 1$. Then $(1,\beta,\beta^2)$ is a basis of $\mathbb{F}_8$ as a vector space over $\mathbb{F}_2$. The elements of $\mathbb{F}_8$ are the eight linear combinations with coefficients 0, 1:

$$(6.12) \qquad \mathbb{F}_8 = \{0,1,\beta,1+\beta,\beta^2,1+\beta^2,\beta+\beta^2,1+\beta+\beta^2\}.$$

Computation in $\mathbb{F}_8$ is done using the relation $\beta^3+\beta+1=0$.

Note that $\mathbb{F}_4$ is not contained in $\mathbb{F}_8$. It couldn't be, because $[\mathbb{F}_8 : \mathbb{F}_2] = 3$, $[\mathbb{F}_4 : \mathbb{F}_2] = 2$, and 2 does not divide 3.

(c) The field $\mathbb{F}_{16}$: The polynomial $x^{16}-x = x(x^{15}-1)$ is divisible in $\mathbb{Z}[x]$ by x^3-1 and by x^5-1. Carrying out the division over the integers gives this factorization:

$$(6.13) \qquad x^{16}-x =$$
$$x(x-1)(x^2+x+1)(x^4+x^3+x^2+x+1)(x^8-x^7+x^5-x^4+x^3-x+1).$$

This is the irreducible factorization in $\mathbb{Z}[x]$. But in $\mathbb{F}_2[x]$, the factor of degree 8 is not irreducible, and

$$(6.14) \qquad x^{16}-x =$$
$$x(x-1)(x^2+x+1)(x^4+x^3+x^2+x+1)(x^4+x^3+1)(x^4+x+1).$$

This factorization displays the three irreducible polynomials of degree 4 in $\mathbb{F}_2[x]$. Note that the factors of x^4-x appear among the factors of $x^{16}-x$. This agrees with the fact that $\mathbb{F}_{16}$ contains $\mathbb{F}_4$.

We will now begin the proof of Theorem (6.4). We will prove the various parts in the following order: (d), (c), (a), (b), (e), and (f).

Proof of Theorem (6.4d). Let K be a field of order q. The multiplicative group $K^\times$ has order $q-1$. Therefore the order of any element $\alpha \in K^\times$ divides $q-1$: $\alpha^{q-1}=1$. This means that α is a root of the polynomial $x^{q-1}-1$. The remaining element of K, zero, is a root of the polynomial x. So every element of K is a root of $x(x^{q-1}-1) = x^q - x$. Since this polynomial has q distinct roots in K, it factors into linear factors in that field:

$$(6.15) \qquad x^q - x = \prod_{\alpha\in K}(x-\alpha).$$

This proves part (d) of the theorem. □

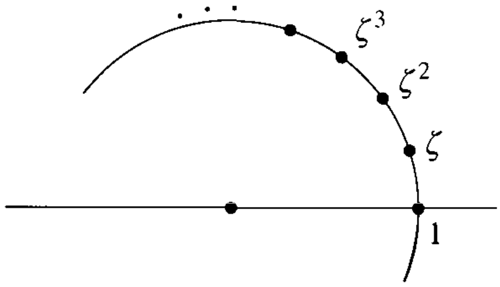
Proof of Theorem (6.4c). By an *n-th root of unity* in a field F, we mean an element α whose nth power is 1. Thus α is an nth root of unity if and only if it is a root of the polynomial

$$(6.16) \qquad x^n - 1,$$

or if and only if its order, as an element of the multiplicative group $F^\times$, divides n. The nonzero elements of a finite field with q elements are $(q-1)$-st roots of unity.

In the field of complex numbers, the nth roots of unity form a cyclic group of order n, generated by

(6.17) $$\zeta_n = e^{2\pi i/n}:$$



A field need not have many roots of unity. For example, the only real ones are ±1. But one property of the complex numbers carries over to arbitrary fields: The nth roots of unity in any field form a cyclic group. For example, in the field $K = \mathbb{F}_4$ of order 4, the group $K^\times$ is a cyclic group of order 3, generated by α. [See (6.3).]

(6.18) **Proposition.** Let F be a field, and let H be a finite subgroup of the multiplicative group $F^\times$, of order n. Then H is a cyclic group, and it consists of all the nth roots of unity in F.

Proof. If H has order n, then the order of an element α of H divides n, so α is an nth root of unity, a root of the polynomial $x^n - 1$. This polynomial has at most n roots, so there aren't any other roots in F [Chapter 11 (1.18)]. It follows that H is the set of all nth roots of unity in F.

It is harder to show that H is cyclic. To do so, we use the Structure Theorem for abelian groups, which tells us that H is isomorphic to a direct product of cyclic groups:

$$H \approx \mathbb{Z}/(d_1) \times \cdots \times \mathbb{Z}/(d_k),$$

where $d_1 | d_2 \cdots | d_k$ and $n = d_1 \cdots d_k$. The order of any element of this product divides d_k because d_k is a common multiple of all the integers d_i. So every element of H is a root of

$$x^{d_k} - 1.$$

This polynomial has at most d_k roots in F. But H contains n elements, and $n = d_1 \cdots d_k$. The only possibility is that $n = d_k$, $k = 1$, and H is cyclic. □

Proof of Theorem (6.4a). We need to prove the existence of a field with q elements. Since we have already proved part (d) of the theorem, we know that the elements of a field of order q are roots of the polynomial $x^q - x$. Also, there exists a field L containing $\mathbb{F}_p$ in which this polynomial (or any given polynomial) factors into linear factors (5.3). The natural thing to try is to take such a field L and hope for the best—that the roots of $x^q - x$ form the subfield K of L we are looking for. This is shown by the following proposition:

(6.19) **Proposition.** Let p be a prime, and let $q = p^r$.

(a) The polynomial $x^q - x$ has no multiple root in any field L of characteristic p.
(b) Let L be a field of characteristic p, and let K be the set of roots of $x^q - x$ in L. Then K is a subfield.

This proposition, combined with Proposition (5.3), proves the existence of a field with q elements.

Proof of Proposition (6.19). (a) The derivative of $x^q - x$ is $qx^{q-1} - 1$. In characteristic p, the coefficient q is equal to 0, so the derivative is equal to -1. Since the constant polynomial -1 has no root, $x^q - x$ and its derivative have no common root! Proposition (5.7) shows that $x^q - x$ has no multiple root.

(b) Let $\alpha, \beta \in L$ be roots of the polynomial $x^q - x$. We have to show that $\alpha \pm \beta$, $\alpha\beta$, and α^{-1} (if $\alpha \neq 0$) are roots of the same polynomial. This is clear for the product and quotient: If $\alpha^q = \alpha$ and $\beta^q = \beta$, then $(\alpha\beta)^q = \alpha\beta$ and $(\alpha^{-1})^q = \alpha^{-1}$. It is not obvious for the sum, and to prove it we use the following proposition:

(6.20) **Proposition.** Let L be a field of characteristic p, and let $q = p^r$. Then in the polynomial ring $L[x,y]$, we have $(x + y)^q = x^q + y^q$.

Proof. We first prove the proposition for the case $q = p$. We expand $(x + y)^p$ in $\mathbb{Z}[x,y]$, obtaining

$$(x + y)^p = x^p + \binom{p}{1}x^{p-1}y + \binom{p}{2}x^{p-2}y^2 + \cdots + \binom{p}{p-1}xy^{p-1} + y^p,$$

by the Binomial Theorem. The binomial coefficient $\binom{p}{r}$ is an integer, and if $0 < r < p$, it is divisible by p [see the proof of (4.6) in Chapter 11]. It follows that the map $\mathbb{Z}[x,y] \longrightarrow L[x,y]$ sends these coefficients to zero and that $(x + y)^p = x^p + y^p$ in $L[x,y]$.

We now treat the general case $q = p^r$ by induction on r: Suppose that the proposition has been proved for integers less than r and that $r > 1$. Let $q' = p^{r-1}$. Then by induction, $(x + y)^q = ((x + y)^{q'})^p = (x^{q'} + y^{q'})^p = (x^{q'})^p + (y^{q'})^p = x^q + y^q$. □

To complete the proof of Proposition (6.19), we evaluate x, y at α, β to conclude that $(\alpha + \beta)^q = \alpha^q + \beta^q$. Then if $\alpha^q = \alpha$ and $\beta^q = \beta$, we find $(\alpha + \beta)^q = \alpha + \beta$, as required. The case of $\alpha - \beta$ follows by substituting $-\beta$ for β. □

Proof of Theorem (6.4b). Let K and K' be fields of order q, and let α be a generator of the cyclic group $K^\times$. Then K is certainly generated as a field extension of $F = \mathbb{F}_p$ by the element α: $K = F(\alpha)$. Let $f(x)$ be the irreducible polynomial for α over F, so that $K \approx F[x]/(f)$ (2.6). Then α is a root of two polynomials: $f(x)$ and $x^q - x$. Since f is irreducible, it divides $x^q - x$ (5.4e). We now go over to the second field K'. Since $x^q - x$ factors into linear factors in K', f has a root α' in K'.

Then $K \approx F[x]/(f) \approx F(\alpha')$. Since K and K' have the same order, $F(\alpha') = K'$; hence K and K' are isomorphic. □

Proof of Theorem (6.4e). Let $f(x)$ be an irreducible polynomial of degree r in $F[x]$, where $F = \mathbb{F}_p$ as before. It has a root α in some field extension L of F, and the subfield $K = F(\alpha)$ of L has degree r over F (3.2). Therefore K has order $q = p^r$, and by part (d) of the theorem, α is also a root of $x^q - x$. Since f is irreducible, it divides $x^q - x$, as required.

In order to prove the same thing for irreducible polynomials whose degree k divides r, it suffices to prove the following lemma:

(6.21) **Lemma.** Let k be an integer dividing r, say $r = ks$, and let $q = p^r$, $q' = p^k$. Then $x^{q'} - x$ divides $x^q - x$.

For if f is irreducible of degree k, then, as above, f divides $x^{q'} - x$, which in turn divides $x^q - x$ in $F[x]$, for any field F.

Proof of the lemma. This is tricky, because we will use the identity

$$y^d - 1 = (y - 1)(y^{d-1} + \cdots + y + 1) \tag{6.22}$$

twice. Substituting $y = q'$ and $d = s$ shows that $q' - 1$ divides $q - 1 = q'^s - 1$. Knowing this, we can conclude that $x^{q'-1} - 1$ divides $x^{q-1} - 1$ by substituting $y = x^{q'-1}$ and $d = (q - 1)/(q' - 1)$. Therefore $x^{q'} - x$ divides $x^q - x$ too. □

So we have shown that every irreducible polynomial whose degree divides r is a factor of $x^q - x$. On the other hand, if f is irreducible and if its degree k doesn't divide r, then since $[K : F] = r$, f doesn't have a root in K, and therefore f doesn't divide $x^q - x$. □

Proof of Theorem (6.4f). If k does not divide r, then $q = p^r$ is not a power of $q' = p^k$, so a field of order q can not be an extension of a field of order q'. On the other hand, if k does divide r, then Lemma (6.21) and part (d) of the theorem show that the polynomial $x^{q'} - x$ has all its roots in a field K of order q. Now Proposition (6.19) shows that K contains a field with q' elements. □

This completes the proof of theorem 6.4.

7. FUNCTION FIELDS

In this section we take a look at *function fields*, the third class of field extensions mentioned in Section 1. The field $\mathbb{C}(x)$ of rational functions in one variable x will be denoted by F throughout the section. Its elements are fractions $g(x) = p(x)/q(x)$ of polynomials $p, q \in \mathbb{C}[x]$, with $q \neq 0$. We usually cancel common factors in p and q so that they have no root in common.

Let us use the symbol P to denote the complex plane, with the complex coordinate x. A rational function $g = p/q$ determines a complex-valued function of x, which is defined for all $x \in P$ such that $q(x) \neq 0$, that is, except at the roots of the polynomial q. Near a root of q, the function defined by g tends to infinity. These roots are called *poles* of g. (We usually use the phrase "rational function" to mean an element of the field of fractions of the polynomial ring. It is unfortunate that the word *function* is already there. This prevents us from modifying the phrase in a natural way when referring to the actual function defined by such a fraction. The terminology is ambiguous, but this can't be helped.)

A minor complication arises because formal rational functions do not define functions at certain points, namely at their poles. When working with the whole field F, we have to face the fact that every value α of x can be a pole of a rational function, for example of the function $(x - \alpha)^{-1}$. There is no way to choose a common domain of definition for all rational functions at once. Fortunately this is not a serious problem, and there are two ways to get around it. One is to introduce an extra value ∞ and to define $g(\alpha) = \infty$ if α is a pole of g. This is actually the better way for many purposes, but for us another way will be easier. It is simply to ignore bad behavior at a finite set of points.

Any particular computations we may make will involve finitely many functions, so they will be valid except at a finite set of points of the plane P, the poles of these functions. A rational function is determined by its value at any infinite set of points. This is proved below, in Lemma (7.2). So we can throw finite sets out of the domain of definition as needed, without losing control of the function. Since a rational function is continuous wherever it is defined, we can recover its value at a point x_0 which was thrown out unnecessarily, as

(7.1)
$$g(x_0) = \lim_{x \to x_0} g(x).$$

(7.2) **Lemma.** If two rational functions f_1, f_2 agree at infinitely many points of the plane, then they are equal elements of F.

Proof. Say that $f_i = p_i/q_i$, where $p_i, q_i \in \mathbb{C}[t]$. Let $h(x) = p_1q_2 - p_2q_1$. If $h(x)$ is the zero polynomial, then $f_1 = f_2$. If $h(x)$ is not zero, then it has finitely many roots, so there are only finitely many points at which $f_1 = f_2$. □

In order to formalize the intuitive procedure of ignoring trouble at finite sets of points, it is convenient to have a notation for the result of throwing out a finite set. Given an infinite set U, we will denote by U' a set obtained from U by deleting an unspecified finite subset, which is allowed to vary as needed:

(7.3)
$$U' = U - (\text{variable finite set}).$$

By a *function* on U' we mean an equivalence class of complex-valued functions, each defined except on a finite subset of U. Two such functions f, g are called *equal on* U' if there is a finite subset Δ of U such that f and g are defined and equal on $U - \Delta$. (We could also refer to this property by saying that $f = g$ *almost everywhere* on U. However, in other contexts, "almost everywhere" often means "except on a set of measure zero," rather than "except on a finite set.") A function f on U' will be called *continuous* if it is represented by a continuous function on some set $U - \Delta$.

The set of continuous functions on U' will be denoted by

$$(7.4) \qquad \mathcal{F}(U) = \{\text{continuous functions on } U'\}.$$

This set forms a ring, with the usual laws of addition and multiplication of functions:

$$(7.5) \qquad [f+g](x) = f(x) + g(x) \quad \text{and} \quad [fg](x) = f(x)g(x).$$

Lemma (7.2) has the following corollary:

(7.6) **Proposition.** The field $F = \mathbb{C}(x)$ is isomorphic to a subring of the ring $\mathcal{F}(P)$, where P is the complex plane. □

Let us now examine one of the simplest function fields in more detail. We are going to need polynomials with coefficients in the field F. Since the symbol x has already been assigned, we use y to denote the new variable. We will study the quadratic field extension K obtained from F by adjoining a root of $f(y)$, where $f = y^2 - x$. Since f depends on the variable x as well as on y, we will also write

$$(7.7) \qquad f = f(x,y) = y^2 - x.$$

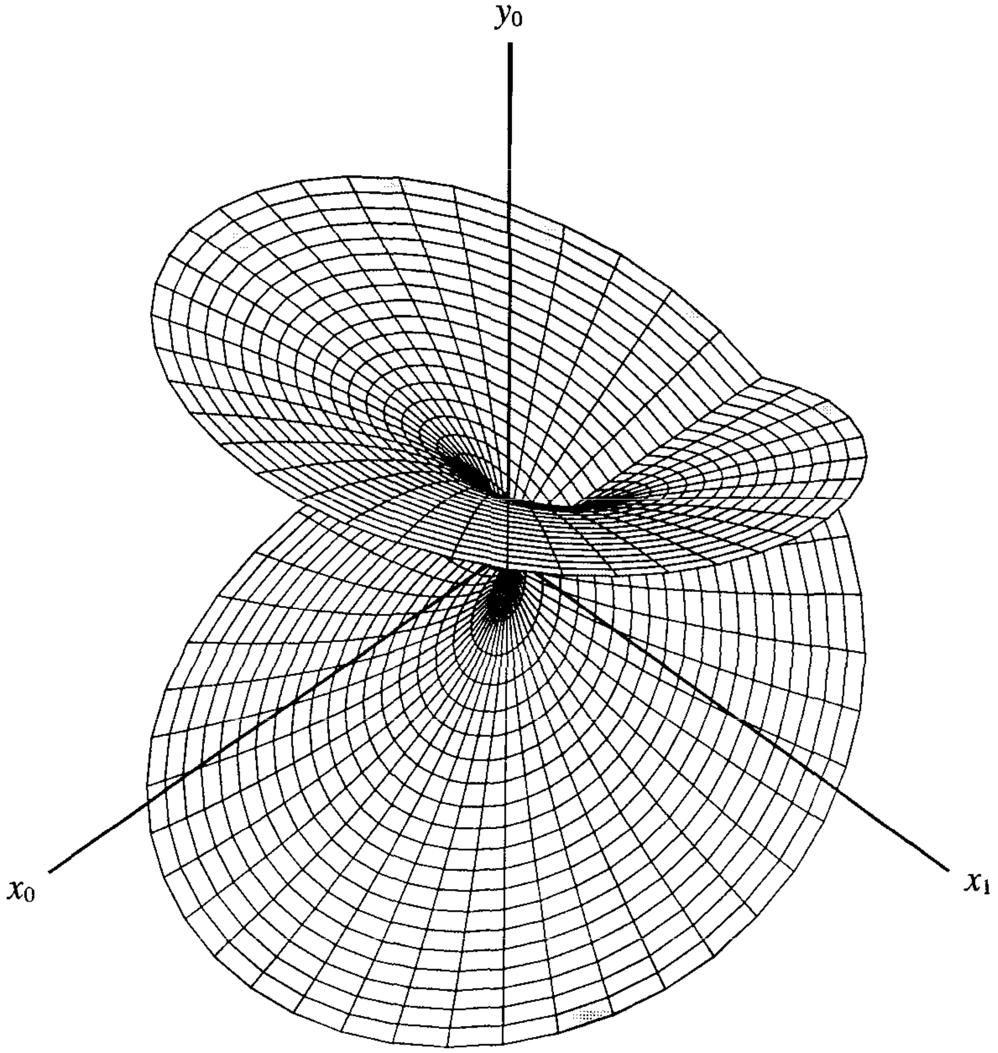
The polynomial $y^2 - x$ is an irreducible element of $F[y]$, so K can be constructed as the abstract field $F[y]/(f)$. The residue of the variable y is a root of f in K.

The importance of function fields comes from the fact that their elements can be interpreted as actual functions. In our case, we can define a square root *function* h, by choosing one of the two values of the square root for each complex number $x : h(x) = \sqrt{x}$. Then h can be interpreted as a function on P'. However, since there are two values of the square root whenever $x \neq 0$, we need to make a lot of choices to define this function. This isn't very satisfactory. If x is real and positive, it is natural to choose the positive square root, but no choice will give a continuous function on the whole complex plane.

The locus S of solutions of the equation $y^2 - x = 0$ in $\mathbb{C}^2$ is called the *Riemann surface* of the polynomial $y^2 - x$ (see Section 8 of Chapter 10). It is depicted below in Figure (7.9), but in order to obtain a surface in real 3-space, we have dropped one coordinate. The complex two-dimensional space $\mathbb{C}^2$ is identified with $\mathbb{R}^4$ by the usual rule $(x,y) = (x_0 + x_1 i, y_0 + y_1 i) \longleftrightarrow (x_0, x_1, y_0, y_1)$. The figure depicts the locus

$$(7.8) \qquad \{(x_0, x_1, y_0) \mid y_0 = \text{real part of } (x_0+x_1 i)^{1/2}\}.$$

This is a projection of S from $\mathbb{R}^4$ to $\mathbb{R}^3$.



(7.9) **Figure.** The Riemann surface $y^2 = x$.

The Riemann surface S does not cut itself along the negative x_0-axis as the projected surface does. Every negative real number x has two purely imaginary square roots, but the real parts of these square roots are zero. This produces the apparent self-crossing in the projected surface. Actually, S is a two-sheeted branched covering of P, as defined in Chapter 10 (8.13), and the only branch point is at $x = 0$.

Figure (7.9) shows the problem encountered when we try to define the square root as a single-valued function. When x is real and positive, the positive square root is the natural choice. We would like to extend this choice continuously over the complex plane, but we run into trouble: Winding once around the origin in complex x-space brings us back to the negative square root. It is better to accept the fact that the square root, as a solution of the equation $y^2 - x = 0$, is a multi-valued function on P'.

Now there is an amazing trick which will allow us to solve any polynomial equation $f(x,y) = 0$ with a *single-valued* function, without making arbitrary choices. The trick is to replace the complex plane P by the Riemann surface S, the locus $f(x,y) = 0$. We are given two functions on S, namely the restrictions of the coordinate functions on $\mathbb{C}^2$. In order to keep things straight, let us introduce new symbols for these functions, say X, Y:

$$(7.10) \qquad \text{X}(x,y) = x \quad \text{and} \quad \text{Y}(x,y) = y, \quad \text{for } (x,y) \in S.$$

These restrictions of the coordinate functions to S are related by the equation $f(\text{X}, \text{Y}) = 0$, because by definition of $S, f(x,y) = 0$ at any point of S.

(7.11) **Proposition.** Let $f(x,y)$ be an irreducible polynomial in $\mathbb{C}[x,y]$ which is not a polynomial in x alone, and let $S = \{(x,y) \mid f(x,y) = 0\}$ be its Riemann surface. Let $K = F[y]/(f)$ be the field extension defined by f. Then K is isomorphic to a subring of the ring $\mathcal{F}(S)$ of continuous functions on S'.

Proof. Let $g(x)$ be a rational function. Since X is the restriction of a coordinate function on $\mathbb{C}^2$, the composed function $g(\text{X})$ is continuous on S except at the points which lie above the poles of g. There are finitely many such points [Chapter 10 (8.11)]. So $g(\text{X})$ is a continuous function on S'. We define a homomorphism $F \longrightarrow \mathcal{F}(S)$ by sending $g(x)$ to $g(\text{X})$. Next, the Substitution Principle extends this map to a homomorphism

$$(7.12) \qquad \varphi \colon F[y] \longrightarrow \mathcal{F}(S),$$

by sending $y \rightsquigarrow \text{Y}$. Since $f(\text{X}, \text{Y}) = 0$, the polynomial $f(x,y)$ is in the kernel of φ. Since $K = F[y]/(f)$, the mapping property of quotients [Chapter 10 (4.2)] gives us a map $\overline{\varphi} \colon K \longrightarrow \mathcal{F}(S)$ which sends the residue of y to Y. Since K is a field, $\overline{\varphi}$ is injective. □

(7.13) **Definition.** An *isomorphism* of branched coverings S_1, S_2 of the plane P is a homeomorphism $\varphi' \colon S_1' \longrightarrow S_2'$ which is compatible with the maps $\pi_i \colon S_i \longrightarrow P$, that is, such that $\pi_2'\varphi = \pi_1'$:

$$\begin{tikzcd} S_1' \arrow[rr, "\varphi'"] \arrow[dr, "\pi_1'"'] & & S_2'. \arrow[dl, "\pi_2'"] \\ & P & \end{tikzcd}$$

By this we mean that φ' is defined except on a finite set of S_1 and that when suitable finite sets are omitted from S_1 and S_2, φ' is a homeomorphism.

A branched covering S is called *connected* if the complement S' of an arbitrary finite set of S is a path-connected set.

We will now state a beautiful theorem which describes the finite extensions of the field of rational functions. Let $\mathcal{E}_n$ denote the set of isomorphism classes of extension fields K of F of degree n. Let $\mathcal{C}_n$ denote the set of isomorphism classes of connected n-sheeted branched coverings $\pi \colon S \longrightarrow P$ of the plane.

(7.14) **Theorem.** *Riemann Existence Theorem:* There is a bijective map $\Phi_n \colon \mathcal{E}_n \longrightarrow \mathcal{C}_n$. If K is the extension obtained by adjoining a root of an irreducible polynomial $f(x,y) \in \mathbb{C}[x,y]$, then the class of branched coverings corresponding to K is represented by the Riemann surface of f. □

The proof of this theorem is a suitable topic for a course in complex variables. It requires too much analysis to give here. Using it, however, we can associate a branched covering of the plane, unique up to isomorphism, to every finite extension field K of F. This covering is called the *Riemann surface of the extension field K.* The Riemann surface of F is the complex plane P itself.

Here are two striking corollaries of the theorem:

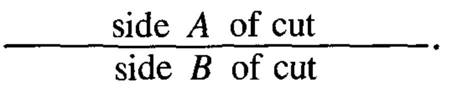
(7.15) **Corollary.** Given a connected n-sheeted branched covering S of the plane, there is a polynomial $f(x,y)$ of degree n in y whose Riemann surface is isomorphic to S.

This follows from the surjectivity of the map Φ_n and from a fact which will be proved in the next chapter [Chapter 14 (4.1)], that every finite extension K of F can be obtained by adjoining a single element. □

(7.16) **Corollary.** Let f, g be irreducible polynomials in $\mathbb{C}[x,y]$, with Riemann surfaces S, T. Let α be a root of $f(y)$ in an extension field of F. If S and T are isomorphic branched coverings, then $g(y)$ has a root in $F(\alpha)$.

This follows from the injectivity of the map Φ_n. □

Visualization of Riemann surfaces is complicated by the fact that they are embedded in $\mathbb{C}^2$, a four-dimensional real space. One aid to constructing and visualizing them is a method known as *cut and paste*. If we cut the surface $y^2 - x$ open along the negative real axis, the double locus in Figure (7.9), then it decomposes into the two parts $\operatorname{re} Y > 0$ and $\operatorname{re} Y < 0$. Each of these parts projects to the x-plane P in a bijective way, if we disregard what happens along the cut. Turning this procedure around, we can construct a surface which is homeomorphic to S in the following way: We stack two copies P_1, P_2 of the complex plane over P and cut them open along the negative real axis $(-\infty, 0]$. These copies of P are called *sheets*. Then we glue side A of P_1 to side B of P_2 and vice versa (see below). Four dimensions are needed to embed S without crossings.

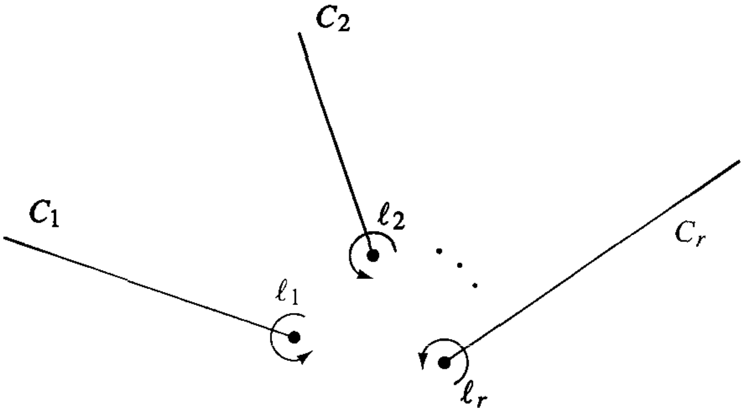


(7.17) **Figure.**

To construct a general branched covering S of the plane by the cut-and-paste procedure, we begin with n copies of the plane P, called *sheets*. The sheets are labelled $P_1, \ldots, P_n$ and are stacked up over P. We also select a finite set of points $\alpha_1, \ldots, \alpha_r$ of P to be branch points. For each branch point α_ν, we choose a curve C_ν

beginning at α_ν and going to infinity in an arbitrary direction. This should be done in such a way that the curves C_ν do not intersect. The sheets P_i are cut open along these curves. Then various sheets are glued to others along opposite edges of the cuts.

To describe the resulting covering S, we need only describe the permutations σ_ν by which the sheets are glued together along the cuts. To be specific, we draw a small loop ℓ_ν around the point α_ν in the counterclockwise direction. Then if the permutation σ_ν sends the index **1** to **3**, we glue sheet P_1 to sheet P_3 as we cross C_ν. This means that if we start on sheet P_1 and wind once around the loop ℓ_ν, we return on sheet P_3. The permutation σ_ν can be arbitrary.



The points α_ν are called *branch points* of the surface S because the surface decomposes into n disjoint sheets near any other point of P. It won't have n disjoint sheets above the point α_ν unless the permutation σ_ν is the identity. If $\sigma_\nu = 1$, then each sheet is glued back to itself along the cut C_ν, so that cut was not needed. But it is convenient to allow this as a possibility. Let's call α_ν a *true* branch point if $\sigma_\nu \neq 1$. Some of the points α_ν may not be true branch points. However, all true branch points are among them.

It is important to note that the numbering of the sheets is arbitrary and, in particular, that the concept of a "top sheet" has no intrinsic meaning for the Riemann surface of a polynomial. If there was a top sheet, we could define y as a single-valued function by choosing the value on that sheet. One can do this only once the Riemann surface has been cut open. This is the whole point; wandering around on the surface will lead us from one sheet to another.

It is not difficult to decide when two such branched coverings are isomorphic.

(7.18) **Proposition.** Let S, T be branched coverings which are constructed as above, with the same branch points α_ν and the same curves C_ν, but using different sets of permutations $(\sigma_1, \ldots, \sigma_r)$ and $(\tau_1, \ldots, \tau_r)$. Then S and T are isomorphic coverings if and only if the two sets of permutations are conjugate, that is, if and only if there is a permutation ρ such that $\tau_\nu = \rho^{-1}\sigma_\nu\rho$ for all ν.

Proof. Let σ, C stand for σ_ν, C_ν. Our rule is that P_i is glued to $P_{i\sigma}$ along C. Suppose that we relabel the sheets $P_1, \ldots, P_n$, changing the numbers by a permutation ρ. To keep old and new labellings straight, let's label the renumbered sheets as Q_j. So for every i, P_i is relabelled as $Q_{i\rho}$. The rule now tells us to glue $P_i = Q_{i\rho}$ to $P_{i\sigma} = Q_{i\sigma\rho}$. Substituting $i = j\rho^{-1}$ shows that the rule glues Q_j to $Q_{j\rho^{-1}\sigma\rho}$. Thus the permutation which describes this gluing rule is the conjugate $\rho^{-1}\sigma_\nu\rho$ of the old permutation σ_ν. Since the covering is not changed by the relabelling process, this shows that a conjugate set of permutations defines an isomorphic covering.

Conversely, let $\varphi: S \longrightarrow T$ be an isomorphism of coverings. Let $P_1, \ldots, P_n$ be the sheets which are used to construct S, and let $Q_1, \ldots, Q_n$ be those used to construct T. Then since P_i is connected and since T, when cut open, is a disjoint union of the open sets Q_j, the image of P_i must be contained in a single sheet Q_j. Since φ is compatible with the projections to P, which are homeomorphisms except on the cuts, the restriction of φ to P_i must be a bijection onto the sheet Q_j. So we can renumber the sheets Q_j so that P_i is mapped to Q_i. This changes the permutations τ_ν to conjugates, as above. So we may assume that φ carries P_i to Q_i. Also, φ is continuous across the cuts. Therefore if crossing the cut C_ν on sheet P_i leads to P_j, then, similarly, crossing on Q_i must lead to Q_j. Therefore $\sigma_\nu = \tau_\nu$. □

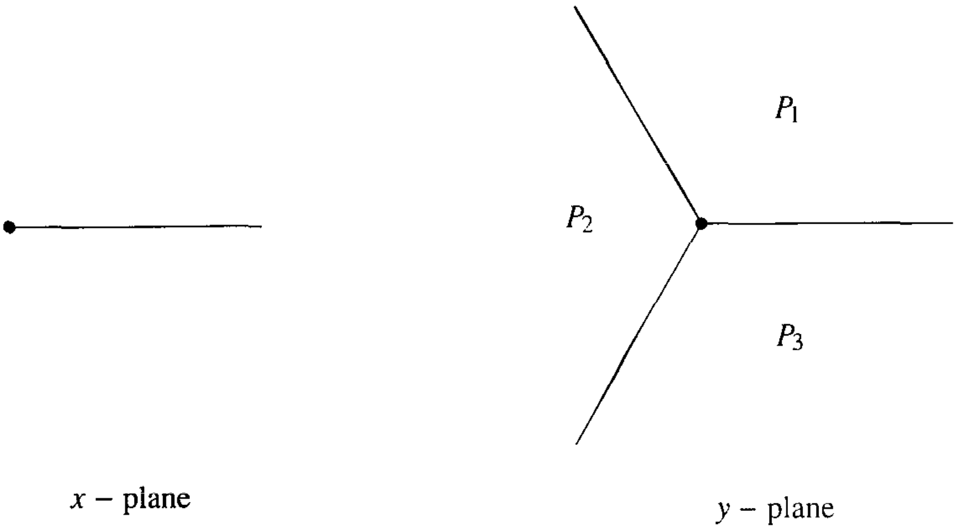
We can also start with an arbitrary branched covering S and reconstruct it in this way: Say that S is branched at the points $\alpha_1, \ldots, \alpha_r \in P$. As above, we choose nonintersecting curves C_i beginning at α_i and going to infinity. Then if S is cut open above the curves C_i, it decomposes into n sheets. This is a theorem of topology, because the complement of the curves C_i in P is simply connected [Munkres, *Topology* p. 342, exc. 8]. Therefore a covering homeomorphic to S can be reconstructed from n sheets $P_1, \ldots, P_n$ by cutting them open along the curves and gluing together to mix up the sheets.

We will now describe the Riemann surfaces of a few simple polynomials f. This is usually difficult to do when f is complicated.

(7.19) **Example.** The Riemann surface of $y^3 - x$: Here y represents a cube root of x, and S is a three-sheeted covering of P. The only branch point is $x = 0$. We cut S open above the positive real axis $C = [0, \infty]$. This decomposes S into three sheets P_1, P_2, P_3, and it is reasonable to guess that the gluing along the cut is done by a cyclic permutation.

This case is fairly easy to analyze because x is a single-valued function of y. Because of this, we can interpret S as the graph of a function from y-space to x-space, which implies that the projection of S onto the complex y-plane is bijective. We identify S with the y-plane using this projection and cut it open above C. This will decompose the plane into three parts corresponding to the sheets P_i. The rules for gluing will be evident when this decomposition is made explicit.

The values of y lying over the cut C are those for which $y^3 = x$ is real and positive. They are $y = re^{i\theta}$, where $\theta = 0$, $2\pi/3$, or $4\pi/3$. So the sheets are sectors.

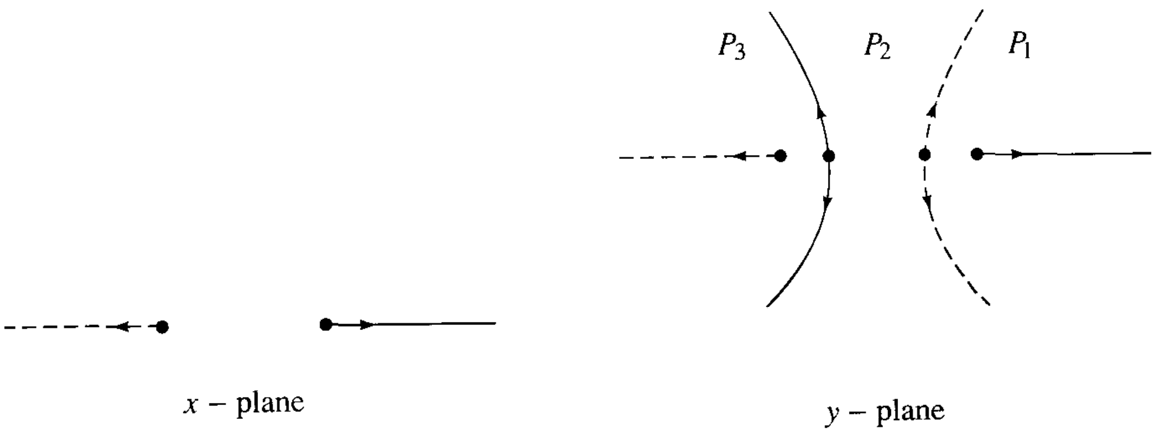


In the figure, the sectors have been numbered arbitrarily. Note that under the map $y \rightsquigarrow y^3 = x$, each of the three sectors is stretched radially and maps bijectively to the entire plane, disregarding the cuts. As we move along S to cross the cut in the x-plane, we also cross one of the three cuts in the y-plane. As predicted, this permutes the sheets by the cyclic permutation $(1\,2\,3)$. □

(7.20) **Example.**

The Riemann surface of $f(x, y) = y^3 - 3y - x$: The points x at which this polynomial has fewer than three roots are found by solving the equations $f = \partial f/\partial y = 0$ [see Chapter 10 (8.12)]. Here $\partial f/\partial y = 3(y^2 - 1)$. So the solutions are $y = \pm 1$, and hence $x = \pm 2$. We may cut S open above the curves $C_1 = (-\infty, -2]$ and $C_2 = [2, \infty)$, to decompose it into three sheets.

Again, x is a single-valued function of y, and we can analyze the gluing of the sheets by cutting the y-plane apart suitably. To do so, we ask for the values of y such that x lies on one of the curves C_i. Since these curves are on the real x-axis, we begin by solving the equation $\operatorname{im} x = 0$. Setting $y = u + vi$, we find $\operatorname{im} x = \operatorname{im}(y^3 - 3y) = v(3u^2 - v^2 - 3)$. The solutions are the u-axis $v = 0$ and the two branches of the hyperbola $3u^2 - v^2 = 3$. The points on the u-axis in the interval $(-2, 2)$ correspond to $x \in (-2, 2)$, so they do not lie over the cuts.

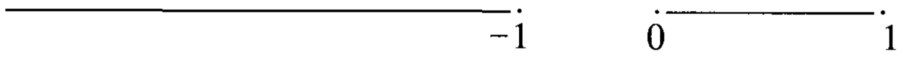


Again, each of the three regions into which the y-plane decomposes is mapped bijectively to the x-plane by the function $y^3 - 3y$, disregarding the cut as always. In the figure, the dotted curves are those which lie over C_1. The figure shows that moving on S to cross the curve $(-\infty, -2]$ interchanges the sheets P_1, P_2, leaving P_3 alone, and similarly that crossing above $[2, \infty)$ interchanges P_2, P_3. So the branching is described by the transposition $(\mathbf{2\,3})$ at the branch point $x = -2$ and by $(\mathbf{1\,2})$ at $x = 2$. □

(7.21) **Example.** The Riemann surface of $y^2 - x^3 + x^2$: There are two points $x = 0, 1$ above which S has fewer than two points. However, at $x = 0$ the sheets cross without getting mixed up, so the only true branch point is $x = 1$. To see this we make the change of variable $x = x$, $z = y/x$, which is defined and invertible except at $x = 0$. Then $z^2 - x + 1 = 0$. The given surface S becomes homeomorphic to the Riemann surface of $z^2 - x + 1$ when the points above the origin are deleted, and the surface can be reduced to (7.9) by a translation in the x-plane. □

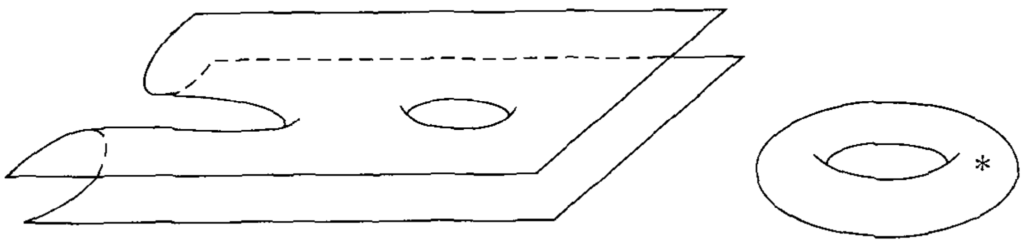
When it is not possible to solve for x as a single-valued function of y, the problem of describing the gluing data becomes more difficult. We will work out one example of this type.

(7.22) **Example.** The Riemann surface of $y^2 - (x^3 - x)$: There are three points at which $x^3 - x = 0$, namely $x = 0, \pm 1$, and the surface has three branch points at which it behaves like the Riemann surface of $y^2 - x$ at the origin. Our systematic procedure is to make cuts from these three branch points to infinity, but in this case another choice of cuts is easier to analyze. The values of x such that y is purely imaginary are the real x such that $x^3 - x \leq 0$. These are the points in the two intervals $(-\infty, -1]$ and $[0, 1]$. If we cut S open along these two intervals, it will decompose into the parts re $y > 0$ and re $y < 0$. Thus we can reconstruct the surface S by stacking up two copies of P, cutting them open along the intervals and gluing to mix up the sheets as before.



(7.23) **Figure.**

The fact that a surface constructed by the cut-and-paste method crosses itself along the cuts makes it confusing to visualize directly. But since the cuts are along the real axis in this example, we can avoid crossings by turning one of the sheets over. This ruins the representation of S as a double covering of P, but the advantage is that the sheets are now glued along the same side of the cut. There are two such cuts in Figure (7.23). Turning one sheet over and stretching to pull the slits apart after gluing results in the following picture: This Riemann surface is homeomorphic to a torus with one point deleted. □



8. TRANSCENDENTAL EXTENSIONS

In this section we will take a brief look at some transcendental field extensions. We saw in Propositon (2.5) that the structure of the field extension $F(\alpha)$ generated by a single transcendental element α over a field F does not depend on the element α. But if two transcendental elements α, β are adjoined at the same time, the structure of the field $F(\alpha, \beta)$ which is obtained will depend on whether or not the elements α and β are algebraically related, and if they are related, the structure will depend on the nature of this relation. For example, $\alpha = \sqrt{\pi}$ and $\beta = \sqrt[4]{\pi}\sqrt{\pi - 1}$ are transcendental numbers over $\mathbb{Q}$, which are related by the equation

$$\beta^2 - \alpha^3 + \alpha = 0.$$

In general, we call a set of elements $\{\alpha_1, \ldots, \alpha_n\}$ of an extension field $K \supset F$ *algebraically dependent over F* if there is a nonzero polynomial in n variables $f(x_1, \ldots, x_n) \in F[x_1, \ldots, x_n]$ such that

$$f(\alpha_1, \ldots, \alpha_n) = 0,$$

and we call them *algebraically independent over F* if there is no such polynomial. Thus $\sqrt{\pi}$ and $\sqrt[4]{\pi}\sqrt{\pi - 1}$ are algebraically dependent over $\mathbb{Q}$. It is conjectured that e and π are algebraically independent, but this has not been proved.

We can interpret algebraic independence in terms of the substitution homomorphism $\varphi: F[x_1, \ldots, x_n] \longrightarrow K$ sending $f(x_1, \ldots, x_n) \rightsquigarrow f(\alpha_1, \ldots, \alpha_n)$. The elements $\alpha_1, \ldots, \alpha_n$ are algebraically independent if $\ker \varphi = 0$, that is, if φ is injective, and algebraically dependent otherwise. Passing to fields of fractions gives this proposition:

(8.1) **Proposition.** If $\alpha_1, \ldots, \alpha_n$ are algebraically independent, then $F(\alpha_1, \ldots, \alpha_n)$ is isomorphic to the field $F(x_1, \ldots, x_n)$ of rational functions in $x_1, \ldots, x_n$, the field of fractions of $F[x_1, \ldots, x_n]$. □

An extension of the form $F(\alpha_1, \ldots, \alpha_n)$, where α_i are algebraically independent, is called a *pure transcendental* extension.

(8.2) **Definition.** A *transcendence basis* for a field extension K of F is a set of elements $(\alpha_1, \ldots, \alpha_n)$ which are algebraically independent and such that K is an algebraic extension of the field $F(\alpha_1, \ldots, \alpha_n)$.

(8.3) **Theorem.** Let $(\alpha_1,\ldots,\alpha_m)$ and $(\beta_1,\ldots,\beta_n)$ be elements in an extension K of a field F. Assume that K is algebraic over $F(\beta_1,\ldots,\beta_n)$ and that $\alpha_1,\ldots,\alpha_m$ are algebraically independent over F. Then $m \leq n$, and $(\alpha_1,\ldots,\alpha_m)$ can be completed to a transcendence basis for K by adding $(n - m)$ of the elements β_i.

We leave the proof of this theorem as an exercise. □

(8.4) **Corollary.** Any two transcendence bases for an extension $F \subset K$ have the same number of elements. □

(8.5) **Definition.** The *transcendence degree* of K is the number of elements in a transcendence basis, or is infinite if no finite transcendence basis exists.

(8.6) **Examples**

(a) The fields $F(x_1,\ldots,x_n)$ of rational functions in n variables are not isomorphic extensions of F for different values of n, because $(x_1,\ldots,x_n)$ is a transcendence basis.

(b) Let α, β be as at the beginning of the section. The single element π forms a transcendence basis for $K = \mathbb{Q}(\alpha,\beta)$ over $\mathbb{Q}$. Therefore (8.3) implies that, as was asserted above, any two elements of K are algebraically dependent. The element β is another transcendence basis.

(c) Consider any two polynomials or rational functions in one variable $f, g \in F(x)$. There is a nonzero polynomial $\varphi(y,z) \in F[y,z]$ such that $\varphi(f,g) = 0$. For, the transcendence degree of $F(x)$ is 1, and hence f, g are algebraically dependent.

Most field extensions aren't pure transcendental, though this may be difficult to decide for a particular extension. Here are two examples:

(8.7) **Proposition.**

(a) The function field $L = \mathbb{C}(x)[y]/(y^2 - x^3)$ is a pure transcendental extension of $\mathbb{C}$. It is the field of rational functions in $t = y/x$.

(b) The function field $K = \mathbb{C}(x)[y]/(y^2 - x^3 + x)$ is not a pure transcendental extension of $\mathbb{C}$. That is, there is no element $t \in K$ such that $K = \mathbb{C}(t)$.

Proof. In both cases, the transcendence degree of K over $\mathbb{C}$ is 1, because x is a transcendence basis.

(a) Let $t = y/x$. Then $\mathbb{C}(t) \subset L$ because $t \in L$. Now L is generated by x and y, by definition. On the other hand, $x = t^2$ and $y = t^3$. Therefore $L = \mathbb{C}(t)$. Since K has transcendence degree 1, (8.4) shows that t is transcendental.

(b) (*Sketch*) To show that K is not a field of rational functions, we appeal to the geometry of its Riemann surface. We saw in the last section that this surface is a torus from which one point has been deleted. On the other hand, the Riemann surface of the field of rational functions $\mathbb{C}(t)$ is the complex plane itself. Now, it is a theorem of topology that the torus and the plane are not homeomorphic and that they do not become homeomorphic when finite sets are deleted. If we admit this theorem, then the next proposition will complete the proof.

(8.8) **Proposition.** Let $K = \mathbb{C}(x)[y]/(f)$ and $L = \mathbb{C}(t)[u]/(g)$ be function fields with Riemann surfaces S, T respectively. A homomorphism $\varphi: L \longrightarrow K$ which is the identity on the subfield $\mathbb{C}$ induces a map $\varphi^*: S' \longrightarrow T$ between their Riemann surfaces, which is defined and continuous except on a finite set of points of S. If φ is an isomorphism, then φ^* becomes a homeomorphism when suitable finite sets are deleted from S and T.

Note that the map φ^* goes from the Riemann surface of K to that of L, in the opposite direction from φ.

Proof. The Riemann surface T is the locus $g(t,u) = 0$ in $\mathbb{C}^2$. According to Proposition (7.11), every element $\alpha \in K$ defines a continuous function on S', so the pair of functions $(\varphi(t), \varphi(u))$ defines a continuous map $S' \longrightarrow \mathbb{C}^2$. Since $g(t,u) = 0$ in L and since φ is a homomorphism which leaves the coefficients of g fixed, $g(\varphi(t), \varphi(u)) = 0$ too. So S' is mapped to T. This is the required map φ^*. If φ is an isomorphism, its inverse defines a map $T' \longrightarrow S$ which is an inverse function to φ^* on the complement of a finite set. □

9. ALGEBRAICALLY CLOSED FIELDS

A field F is said to be *algebraically closed* if every polynomial $f(x) \in F[x]$ of positive degree has a root in F. The fact that the field $\mathbb{C}$ of complex numbers is algebraically closed is called the Fundamental Theorem of Algebra.

(9.1) **Theorem.** *Fundamental Theorem of Algebra*: Every nonconstant polynomial with complex coefficients has a complex root.

We have used this theorem often already. A proof is at the end of the section.

If a field F is algebraically closed, then every nonconstant polynomial $f(x) \in F[x]$ has a linear factor $x - \alpha$, so the only irreducible polynomials are the linear ones. Consequently every polynomial is a product of linear factors. Also, there are no algebraic extensions of F other than F itself (whence the phrase algebraically closed). For if α is algebraic over F, then it is a root of a monic irreducible polynomial $f(x) \in F[x]$. This polynomial must have the form $x - \alpha$, so $\alpha \in F$.

It may be convenient to think of a field F which is being studied as a subfield of an algebraically closed field. For instance, we like to think of number fields as subfields of $\mathbb{C}$. Let us call an extension field K of F an *algebraic closure of F* if

(9.2) (i) K is algebraic over F, and
(ii) K is algebraically closed.

(9.3) **Corollary.** Let F be a subfield of $\mathbb{C}$. The subset $\overline{F}$ of $\mathbb{C}$ consisting of all numbers which are algebraic over F is an algebraic closure of F.

Proof. The fact that $\overline{F}$ is a field has been proved (3.10). To show that $\overline{F}$ is algebraically closed, let $f(x) \in \overline{F}[x]$ be a nonconstant polynomial. Then $f(x)$ has a root α in $\mathbb{C}$, and $\overline{F}(\alpha)$ is algebraic over $\overline{F}$. Since $\overline{F}$ is algebraic over F, α is algebraic over F by (3.11). So $\alpha \in \overline{F}$. □

It is not hard to construct an algebraic closure of a finite field $\mathbb{F}_p$, as a union of the fields $\mathbb{F}_q$, where $q = p^r$ is a power of p. To do this, we choose a sequence of integers $r_1, r_2, \ldots$ with these properties: (i) r_i divides r_{i+1}, and (ii) every integer n divides some r_i. We may take $r_i = i!$, for example. We set $q_i = p^{r_i}$ and $F_i = \mathbb{F}_{q_i}$. It follows from (i) that F_{i+1} contains a subfield isomorphic to F_i (6.4), so we can build a tower of fields $F_1 \subset F_2 \subset \cdots$. Let $\overline{F}$ be the union of this chain of fields. Then (ii) tells us that every finite field $\mathbb{F}_q$, $q = p^r$, is isomorphic to a subfield of some F_i, hence to a subfield of $\overline{F}$. This field is an algebraic closure of $\mathbb{F}_p$.

The following theorem can be proved using Zorn's Lemma.

(9.4) **Theorem.** Every field F has an algebraic closure, and if K_1, K_2 are two algebraic closures of F, there is an isomorphism $\varphi: K_1 \longrightarrow K_2$ which is the identity map on the subfield F. □

Thus the algebraic closure is essentially unique.

(9.5) **Corollary.** Let $\overline{F}$ be an algebraic closure of F, and let K be any algebraic extension of F. There is a subextension $K' \subset \overline{F}$ isomorphic to K. □

Proof of the Fundamental Theorem of Algebra. To show that $f(x_0) = 0$, it is enough to show that the absolute value $|f(x_0)|$ is zero. The existence of such a value $x_0 \in \mathbb{C}$ is proved by the following two lemmas:

(9.6) **Lemma.** Let $f(x)$ be a nonconstant polynomial, and let $x_0 \in \mathbb{C}$ be a point at which $f(x_0) \neq 0$. Then $|f(x_0)|$ is not the minimum value of $|f(x)|$.

(9.7) **Lemma.** Let $f(x)$ be a complex polynomial. Then $|f(x)|$ takes on a minimum value at some point $x_0 \in \mathbb{C}$.

Proof of Lemma (9.6). We first note that the polynomial $x^k - c$ has a root for all $c \in \mathbb{C}$. A nonnegative real number r has a real kth root because the continuous function x^k, which is zero when $x = 0$ and large when x is a large real number, takes on all real values ≥ 0, by the Intermediate Value Theorem. We write the complex number c in the form $c = re^{i\theta}$, where $r = |c|$ and $\theta = \arg c$. Let s be a real kth root of r. Then the required kth root of c is

$$\alpha = se^{i\theta/k}. \tag{9.8}$$

Now let $f(x)$ be a nonconstant polynomial, and let $x_0 \in \mathbb{C}$ be a point at which $f(x_0) \neq 0$. It is convenient to normalize f. We make a change of variable, replacing

x by $x + x_0$, to shift the point in question to the origin: $x_0 = 0$. We also multiply $f(x)$ by $f(0)^{-1}$. Then $f(0) = 1$, and we must show that 1 is not the minimum value of $|f(x)|$.

Let k denote the lowest nonzero power of x occurring in f, so that

$$f(x) = 1 + ax^k + (\text{terms of degree} > k).$$

Let α be a kth root of $-a^{-1}$. We make a final change of variable, replacing x by αx. Then f takes the form

$$f(x) = 1 - x^k + (\text{higher-degree terms}) = 1 - x^k + x^{k+1}g(x),$$

for some polynomial $g(x)$. For small positive real x, the Triangle Inequality shows that

$$|f(x)| \leq |1 - x^k| + |x^{k+1}g(x)| = 1 - x^k + x^{k+1}|g(x)| = 1 - x^k(1 - x|g(x)|).$$

Since $x|g(x)|$ is small for small x, the term $x^k(1 - x|g(x)|)$ is positive when x is a sufficiently small positive real number. For such x, $|f(x)| < |f(0)|$. □

Proof of Lemma (9.7). We may assume that f is not a constant polynomial. For large x, $f(x)$ is also large:

(9.9) $$|f(x)| \longrightarrow \infty \text{ as } |x| \longrightarrow \infty.$$

To prove this, the constant term of f is irrelevant, so we may suppose that it is zero. Then $f(x)$ is divisible by x: $f(x) = xg(x)$. By induction on the degree, the assertion is true for $g(x)$, or else $g(x)$ is constant, and it follows for $f(x)$ as well.

Now since $f(x)$ is large for large x, the greatest lower bound m of $|f(x)|$ in the whole complex plane is also the greatest lower bound in a sufficiently large disc $|x| \leq r$. Since the disc is compact and $|f(x)|$ is a continuous function, it takes on a minimum value in the disc. □

There are several other proofs of the Fundamental Theorem of Algebra, and one of them is particularly appealing, though it is not as easy to make precise as the one just given. We will present it in outline. As before, our problem is to prove that a nonconstant polynomial

(9.10) $$f(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1z + a_0$$

has a root. If $a_0 = 0$, then 0 is a root, so we may assume that $a_0 \neq 0$. We consider the function f: $\mathbb{C} \longrightarrow \mathbb{C}$ defined by the polynomial (9.10).

Let C_r denote a circle of radius r about the origin. We study the images $f(C_r)$ of the circle C_r. To do this, we use polar coordinates, writing $z = re^{i\theta}$. Then $z^n = r^ne^{in\theta}$. As θ runs from 0 to 2π, the point z winds once around the circle of radius r. At the same time, $n\theta$ runs from 0 to $2\pi n$, so the point z^n winds n times around the circle of radius r^n.

For sufficiently large r, the term z^n is dominant in the expression (9.10), and we will have

$$|f(z) - z^n| \leq \tfrac{1}{2}r^n.$$

The proof of this fact is similar to the proof of Lemma (9.6). For our purposes, the factor $\frac{1}{2}$ could be replaced by any positive real number less than 1. This inequality shows us that, as z^n winds n times around the circle of radius r^n, $f(z)$ also winds n times around the origin. A good way to visualize this conclusion is with the dog-on-a-leash model. If someone walks a dog n times around the block, the dog also goes around n times, though following a different path. This will be true provided that the leash is shorter than the radius of the block. Here z^n represents the position of the person at the time θ, and $f(z)$ represents the position of the dog. The length of the leash is $\frac{1}{2}r^n$.

We now vary the radius r. Since f is a continuous function, the image $f(C_r)$ will vary continuously with r. When the radius r is very small, $f(C_r)$ makes a small loop around the constant term a_0 of f. This small loop won't wind around the origin at all. But as we just saw, $f(C_r)$ winds n times around the origin if r is large enough. The only explanation for this is that for some intermediate radius r', $f(C_{r'})$ passes through the origin. This means that for some point α on the circle $C_{r'}$, $f(\alpha) = 0$. This number α is a root of f.

Note that all n loops have to cross the origin, which agrees with the fact that a polynomial of degree n has n roots.

I don't consider this algebra,
but this doesn't mean that algebraists can't do it.

Garrett Birkhoff

EXERCISES

1. Examples of Fields

1. Let F be a field. Find all elements $a \in F$ such that $a = a^{-1}$.
2. Let K be a subfield of $\mathbb{C}$ which is not contained in $\mathbb{R}$. Prove that K is a dense subset of $\mathbb{C}$.
3. Let R be an integral domain containing a field F as subring and which is finite-dimensional when viewed as vector space over F. Prove that R is a field.
4. Let F be a field containing exactly eight elements. Prove or disprove: The characteristic of F is 2.

2. Algebraic and Transcendental Elements

1. Let α be the real cube root of 2. Compute the irreducible polynomial for $1 + \alpha^2$ over $\mathbb{Q}$.
2. Prove Lemma (2.7), that $(1, \alpha, \alpha^2, \ldots, \alpha^{n-1})$ is a basis of $F[\alpha]$.
3. Determine the irreducible polynomial for $\alpha = \sqrt{3} + \sqrt{5}$ over each of the following fields.
(a) $\mathbb{Q}$ (b) $\mathbb{Q}(\sqrt{5})$ (c) $\mathbb{Q}(\sqrt{10})$ (d) $\mathbb{Q}(\sqrt{15})$

4. Let α be a complex root of the irreducible polynomial $x^3 - 3x + 4$. Find the inverse of $\alpha^2 + \alpha + 1$ in $F(\alpha)$ explicitly, in the form $a + b\alpha + c\alpha^2$, $a, b, c \in \mathbb{Q}$.

5. Let $K = F(\alpha)$, where α is a root of the irreducible polynomial $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$. Determine the element α^{-1} explicitly in terms of α and of the coefficients a_i.

6. Let $\beta = \zeta\sqrt[3]{2}$, where $\zeta = e^{2\pi i/3}$, and let $K = \mathbb{Q}(\beta)$. Prove that -1 can not be written as a sum of squares in K.

3. The Degree of a Field Extension

1. Let F be a field, and let α be an element which generates a field extension of F of degree 5. Prove that α^2 generates the same extension.

2. Let $\zeta = e^{2\pi i/7}$, and let $\eta = e^{2\pi i/5}$. Prove that $\eta \notin \mathbb{Q}(\zeta)$.

3. Define $\zeta_n = e^{2\pi i/n}$. Find the irreducible polynomial over $\mathbb{Q}$ of **(a)** ζ_4, **(b)** ζ_6, **(c)** ζ_8, **(d)** ζ_9, **(e)** ζ_{10}, **(f)** ζ_{12}.

4. Let $\zeta_n = e^{2\pi i/n}$. Determine the irreducible polynomial over $\mathbb{Q}(\zeta_3)$ of **(a)** ζ_6, **(b)** ζ_9, **(c)** ζ_{12}.

5. Prove that an extension K of F of degree 1 is equal to F.

6. Let a be a positive rational number which is not a square in $\mathbb{Q}$. Prove that $\sqrt[4]{a}$ has degree 4 over $\mathbb{Q}$.

7. Decide whether or not i is in the field **(a)** $\mathbb{Q}(\sqrt{-2})$, **(b)** $\mathbb{Q}(\sqrt[4]{-2})$, **(c)** $\mathbb{Q}(\alpha)$, where $\alpha^3 + \alpha + 1 = 0$.

8. Let K be a field generated over F by two elements α, β of relatively prime degrees m, n respectively. Prove that $[K:F] = mn$.

9. Let α, β be complex numbers of degree 3 over $\mathbb{Q}$, and let $K = \mathbb{Q}(\alpha, \beta)$. Determine the possibilities for $[K:\mathbb{Q}]$.

10. Let α, β be complex numbers. Prove that if $\alpha + \beta$ and $\alpha\beta$ are algebraic numbers, then α and β are also algebraic.

11. Let α, β be complex roots of irreducible polynomials $f(x), g(x) \in \mathbb{Q}[x]$. Let $F = \mathbb{Q}[\alpha]$ and $K = \mathbb{Q}[\beta]$. Prove that $f(x)$ is irreducible in K if and only if $g(x)$ is irreducible in F.

12. **(a)** Let $F \subset F' \subset K$ be field extensions. Prove that if $[K:F] = [K:F']$, then $F = F'$.
(b) Give an example showing that this need not be the case if F is not contained in F'.

13. Let $\alpha_1, \ldots, \alpha_k$ be elements of an extension field K of F, and assume that they are all algebraic over F. Prove that $F(\alpha_1, \ldots, \alpha_k) = F[\alpha_1, \ldots, \alpha_k]$.

14. Prove or disprove: Let α, β be elements which are algebraic over a field F, of degrees d, e respectively. The monomials $\alpha^i\beta^j$ with $i = 0, \ldots, d - 1$, $j = 0, \ldots, e - 1$ form a basis of $F(\alpha, \beta)$ over F.

15. Prove or disprove: Every algebraic extension is a finite extension.

4. Constructions with Ruler and Compass

1. Express cos 15° in terms of square roots.

2. Prove that the regular pentagon can be constructed by ruler and compass **(a)** by field theory, and **(b)** by finding an explicit construction.

3. Derive formula (4.12).
4. Determine whether or not the regular 9-gon is constructible by ruler and compass.
5. Is it possible to construct a square whose area is equal to that of a given triangle?
6. Let α be a real root of the polynomial $x^3 + 3x + 1$. Prove that α can not be constructed by ruler and compass.
7. Given that π is a transcendental number, prove the impossibility of squaring the circle by ruler and compass. (This means constructing a square whose area is the same as the area of a circle of unit radius.)
8. Prove the impossibility of "duplicating the cube," that is, of constructing the side length of a cube whose volume is 2.
9. (a) Referring to the proof of Proposition (4.8), prove that the discriminant D is negative if and only if the circles do not intersect.
   (b) Determine the line which appears at the end of the proof of Proposition (4.8) geometrically if $D \geq 0$ and also if $D < 0$.
10. Prove that if a prime integer p has the form $2^r + 1$, then it actually has the form $2^{2^k} + 1$.
11. Let C denote the field of constructible real numbers. Prove that C is the smallest subfield of $\mathbb{R}$ with the property that if $a \in C$ and $a > 0$, then $\sqrt{a} \in C$.
12. The points in the plane can be considered as complex numbers. Describe the set of constructible points explicitly as a subset of $\mathbb{C}$.
13. Characterize the constructible real numbers in the case that three points are given in the plane to start with.

*14. Let the rule for construction in three-dimensional space be as follows:
   (i) Three non-collinear points are given. They are considered to be constructed.
   (ii) One may construct a plane through three non-collinear constructed points.
   (iii) One may construct a sphere with center at a constructed point and passing through another constructed point.
   (iv) Points of intersection of constructed planes and spheres are considered to be constructed if they are isolated points, that is, if they are not part of an intersection curve.

   Prove that one can introduce coordinates, and characterize the coordinates of the constructible points.

5. *Symbolic Adjunction of Roots*

1. Let F be a field of characteristic zero, let f' denote the derivative of a polynomial $f \in F[x]$, and let g be an irreducible polynomial which is a common divisor of f and f'. Prove that g^2 divides f.
2. For which fields F and which primes p does $x^p - x$ have a multiple root?
3. Let F be a field of characteristic p.
   (a) Apply (5.7) to the polynomial $x^p + 1$.
   (b) Factor this polynomial into irreducible factors in $F[x]$.
4. Let $\alpha_1,\dots,\alpha_n$ be the roots of a polynomial $f \in F[x]$ of degree n in an extension field K. Find the best upper bound that you can for $[F(\alpha_1,\dots,\alpha_n) : F]$.

6. Finite Fields

1. Identify the group $\mathbb{F}_4{}^+$.
2. Write out the addition and multiplication tables for $\mathbb{F}_4$ and for $\mathbb{Z}/(4)$, and compare them.
3. Find a thirteenth root of 3 in the field $\mathbb{F}_{13}$.
4. Determine the irreducible polynomial over $\mathbb{F}_2$ for each of the elements (6.12) of $\mathbb{F}_8$.
5. Determine the number of irreducible polynomials of degree 3 over the field $\mathbb{F}_3$.
6. **(a)** Verify that (6.9, 6.10, 6.13) are irreducible factorizations over $\mathbb{F}_2$.
   (b) Verify that (6.11, 6.13) are irreducible factorizations over $\mathbb{Z}$.
7. Factor $x^9 - x$ and $x^{27} - x$ in $\mathbb{F}_3$. Prove that your factorizations are irreducible.
8. Factor the polynomial $x^{16} - x$ in the fields **(a)** $\mathbb{F}_4$ and **(b)** $\mathbb{F}_8$.
9. Determine all polynomials $f(x)$ in $\mathbb{F}_q[x]$ such that $f(\alpha) = 0$ for all $\alpha \in \mathbb{F}_q$.
10. Let K be a finite field. Prove that the product of the nonzero elements of K is -1.
11. Prove that every element of $\mathbb{F}_p$ has exactly one pth root.
12. Complete the proof of Proposition (6.19) by showing that the difference $\alpha - \beta$ of two roots of $x^q - x$ is a root of the same polynomial.
13. Let p be a prime. Describe the integers n such that there exist a finite field K of order n and an element $\alpha \in K^\times$ whose order in $K^\times$ is p.
14. Work this problem without appealing to Theorem (6.4).
    (a) Let $F = \mathbb{F}_p$. Determine the number of monic irreducible polynomials of degree 2 in $F[x]$.
    (b) Let $f(x)$ be one of the polynomials described in (a). Prove that $K = F[x]/(f)$ is a field containing p^2 elements and that the elements of K have the form $a + b\alpha$, where $a, b \in F$ and α is a root of f in K. Show that every such element $a + b\alpha$ with $b \neq 0$ is the root of an irreducible quadratic polynomial in $F[x]$.
    (c) Show that every polynomial of degree 2 in $F[x]$ has a root in K.
    (d) Show that all the fields K constructed as above for a given prime p are isomorphic.
15. The polynomials $f(x) = x^3 + x + 1$, $g(x) = x^3 + x^2 + 1$ are irreducible over $\mathbb{F}_2$. Let K be the field extension obtained by adjoining a root of f, and let L be the extension obtained by adjoining a root of g. Describe explicitly an isomorphism from K to L.
16. **(a)** Prove Lemma (6.21) for the case $F = \mathbb{C}$ by looking at the roots of the two polynomials.
    (b) Use the principle of permanence of identities to derive the conclusion when F is an arbitrary ring.

7. Function Fields

1. Determine a real polynomial in three variables whose locus of zeros is the projected Riemann surface (7.9).
2. Prove that the set $\mathcal{F}(U)$ of continuous functions on U' forms a ring.
3. Let $f(x)$ be a polynomial in $F[x]$, where F is a field. Prove that if there is a rational function $r(x)$ such that $r^2 = f$, then r is a polynomial.
4. Referring to the proof of Proposition (7.11), explain why the map $F \longrightarrow \mathcal{F}(S)$ defined by $g(x) \rightsquigarrow g(\mathrm{X})$ is a homomorphism.

5. Determine the branch points and the gluing data for the Riemann surfaces of the following polynomials.
   (a) $y^2 - x^2 + 1$ **(b)** $y^5 - x$ **(c)** $y^4 - x - 1$ **(d)** $y^3 - xy - x$
   (e) $y^3 - y^2 - x$ **(f)** $y^3 - x(x-1)$ **(g)** $y^3 - x(x-1)^2$ **(h)** $y^3 + xy^2 + x$
   (i) $x^2y^2 - xy - x$

6. **(a)** Determine the number of isomorphism classes of function fields K of degree 3 over $F = \mathbb{C}(x)$ which are ramified only at the points ± 1.
   (b) Describe the gluing data for the Riemann surface corresponding to each isomorphism class of fields as a pair of permutations.
   (c) For each isomorphism class, determine a polynomial $f(x, y)$ such that $K = F[x]/(f)$ represents the isomorphism class.

*7. Prove the Riemann Existence Theorem for quadratic extensions.

*8. Let S be a branched covering constructed with branch points $\alpha_1, \ldots, \alpha_r$, curves $C_1, \ldots, C_r$, and permutations $\sigma_1, \ldots, \sigma_r$. Prove that S is connected if and only if the subgroup Σ of the symmetric group S_n which is generated by the permutations σ_ν operates transitively on the indices **1**,..., **n**.

*9. It can be shown that the Riemann surface S of a function field is homeomorphic to the complement of a finite set of points in a compact oriented two-dimensional manifold $\overline{S}$. The *genus* of such a surface is defined to be the number of holes in the corresponding manifold $\overline{S}$. So if $\overline{S}$ is a sphere, the genus of S is 0, while if $\overline{S}$ is a torus, the genus of S is 1. The genus of a function field is defined to be the genus of its Riemann surface. Determine the genus of the field defined by each polynomial.
   (a) $y^2 - (x^2 - 1)(x^2 - 4)$ **(b)** $y^2 - x(x^2 - 1)(x^2 - 4)$ **(c)** $y^3 + y + x$
   (d) $y^3 - x(x-1)$ **(e)** $y^3 - x(x-1)^2$

8. Transcendental Extensions

1. Let $K = F(\alpha)$ be a field extension generated by an element α, and let $\beta \in K$, $\beta \notin F$. Prove that α is algebraic over the field $F(\beta)$.
2. Prove that the isomorphism $\mathbb{Q}(\pi) \longrightarrow \mathbb{Q}(e)$ sending $\pi \rightsquigarrow e$ is discontinuous.
3. Let $F \subset K \subset L$ be fields. Prove that $\operatorname{tr\,deg}_F L = \operatorname{tr\,deg}_F K + \operatorname{tr\,deg}_K L$.
4. Let $(\alpha_1, \ldots, \alpha_n) \subset K$ be an algebraically independent set over F. Prove that an element $\beta \in K$ is transcendental over $F(\alpha_1, \ldots, \alpha_n)$ if and only if $(\alpha_1, \ldots, \alpha_n; \beta)$ is algebraically independent.
5. Prove Theorem (8.3).

9. Algebraically Closed Fields

1. Derive Corollary (9.5) from Theorem (9.4).
2. Prove that the field $\overline{F}$ constructed in this text as the union of finite fields is algebraically closed.

*3. With notation as at the end of the section, a comparison of the images $f(C_r)$ for varying radii shows another interesting geometric feature: For large r, the curve $f(C_r)$ has n loops. This can be expressed formally by saying that its total curvature is $2\pi n$. For small r, the linear term $a_1 z + a_0$ dominates $f(z)$. Then $f(C_r)$ makes a single loop around a_0. Its total curvature is only 2π. Something happens to the loops and the curvature, as r varies. Explain.

***4.** If you have access to a computer with a good graphics system, use it to illustrate the variation of $f(C_r)$ with r. Use log-polar coordinates ($\log r$, θ).

Miscellaneous Exercises

1. Let $f(x)$ be an irreducible polynomial of degree 6 over a field F, and let K be a quadratic extension of F. Prove or disprove: Either f is irreducible over K, or else f is a product of two irreducible cubic polynomials over K.

2. **(a)** Let p be an odd prime. Prove that exactly half of the elements of $\mathbb{F}_p^\times$ are squares and that if α, β are nonsquares, then $\alpha\beta$ is a square.

(b) Prove the same as (a) for any finite field of odd order.

(c) Prove that in a finite field of even order, every element is a square.

3. Write down the irreducible polynomial for $\alpha = \sqrt{2} + \sqrt{3}$ over $\mathbb{Q}$ and prove that it is reducible modulo p for every prime p.

***4.** **(a)** Prove that any element of $GL_2(\mathbb{Z})$ of finite order has order $1, 2, 3, 4$, or 6.

(b) Extend this theorem to $GL_3(\mathbb{Z})$, and show that it fails in $GL_4(\mathbb{Z})$.

5. Let c be a real number, not ± 2. The plane curve C: $x^2 + cxy + y^2 = 1$ can be parametrized rationally. To do this, we choose the point $(0, 1)$ on C and parametrize the lines through this point by their slope: L_t: $y = tx + 1$. The point at which the line L_t intersects C can be found algebraically.

(a) Find the equation of this point explicitly.

(b) Use this procedure to find all solutions of the equation $x^2 + cxy + y^2 = 1$ in the field $F = \mathbb{F}_p$, when c is in that field and $c \neq \pm 2$.

(c) Show that the number of solutions is $p - 1$, p, or $p + 1$, and describe how this number depends on the roots of the polynomial $t^2 + ct + 1$.

6. The *degree* of a rational function $f(x) = p(x)/q(x) \in \mathbb{C}(x)$ is defined to be the maximum of the degrees of p and q, when p, q are chosen to be relatively prime. Every rational function f defines a map $P' \longrightarrow P'$, by $x \rightsquigarrow f(x)$. We will denote this map by f too.

(a) Suppose that f has degree d. Show that for any point y_0 in the plane, the fibre $f^{-1}(y_0)$ contains at most d points.

(b) Show that $f^{-1}(y_0)$ consists of precisely d points, except for a finite number of y_0. Identify the values y_0 where there are fewer than d points in terms of f and df/dx.

***7.** **(a)** Prove that a rational function $f(x)$ generates the field of rational functions $\mathbb{C}(x)$ if and only if it is of the form $(ax + b)/(cx + d)$, with $ad - bc \neq 0$.

(b) Identify the group of automorphisms of $\mathbb{C}(x)$ which are the identity on $\mathbb{C}$.

***8.** Let K/F be an extension of degree 2 of rational function fields, say $K = \mathbb{C}(t)$ and $F = \mathbb{C}(x)$. Prove that there are generators x', t' for the two fields, such that $t = (\alpha t' + \beta)/(\gamma t' + \delta)$ and $x = (ax' + b)/(cx' + d)$, $\alpha, \beta, \gamma, \delta, a, b, c, d \in \mathbb{C}$, such that $t'^2 = x'$.

***9.** Fill in the following outline to give an algebraic proof of the fact that $K = \mathbb{C}(x)[y]/(y^2 - x^3 + x)$ is not a pure transcendental extension of $\mathbb{C}$. Suppose that $K = \mathbb{C}(t)$ for some t. Then x and y are rational functions of t.

(a) Using the result of the previous problem and replacing t by t' as necessary, reduce to the case that $x = (at^2 + b)/(ct^2 + d)$.

(b) Say that $y = p(t)/q(t)$. Then the equation $y^2 = x(x + 1)(x - 1)$ reads

$$\frac{p(t)^2}{q(t)^2} = \frac{(at^2 + b)((a + c)t^2 + b + d)((a - c)t^2 + b - d)}{(ct^2 + d)^3}.$$

Either the numerators and denominators on the two sides agree, or else there is cancellation on the right side.

(c) Complete the proof by analyzing the two possibilities given in (b).

***10.** **(a)** Prove that the homomorphism $SL_2(\mathbb{Z}) \longrightarrow SL_2(\mathbb{F}_p)$ obtained by reducing the matrix entries modulo 2 is surjective.

(b) Prove the analogous assertion for SL_n.

***11.** Determine the conjugacy classes of elements order 2 in $GL_n(\mathbb{Z})$.