such that $g(y_i) = x_i$ for all i, and $f \circ g$ and $g \circ f$ are the respective identity mappings.

Corollary 4.3. *Two modules having bases whose cardinalities are equal are isomorphic.*

Proof. Clear.

We shall leave the proofs of the following statements as exercises.

Let M be a free module over A, with basis $\{x_i\}_{i \in I}$, so that

$$M = \bigoplus_{i \in I} Ax_i.$$

Let $\mathfrak{a}$ be a two sided ideal of A. Then $\mathfrak{a}M$ is a submodule of M. Each $\mathfrak{a}x_i$ is a submodule of Ax_i. *We have an isomorphism* (of A-modules)

$$\boxed{M/\mathfrak{a}M \approx \bigoplus_{i \in I} Ax_i/\mathfrak{a}x_i.}$$

Furthermore, each $Ax_i/\mathfrak{a}x_i$ is isomorphic to $A/\mathfrak{a}$, as A-module.

Suppose in addition that A is commutative. Then $A/\mathfrak{a}$ is a ring. Furthermore $M/\mathfrak{a}M$ is a free module over $A/\mathfrak{a}$, and each $Ax_i/\mathfrak{a}x_i$ is free over $A/\mathfrak{a}$. If $\bar{x}_i$ is the image of x_i under the canonical homomorphism

$$Ax_i \to Ax_i/\mathfrak{a}x_i,$$

then the single element $\bar{x}_i$ is a basis of $Ax_i/\mathfrak{a}x_i$ over $A/\mathfrak{a}$.

All of these statements should be easily verified by the reader. Now let A be an arbitrary commutative ring. A module M is called **principal** if there exists an element $x \in M$ such that $M = Ax$. The map

$$a \mapsto ax \text{ (for } a \in A)$$

is an A-module homomorphism of A onto M, whose kernel is a left ideal $\mathfrak{a}$, and inducing an isomorphism of A-modules

$$A/\mathfrak{a} \approx M.$$

Let M be a finitely generated module, with generators $\{v_1, \ldots, v_n\}$. Let F be a free module with basis $\{e_1, \ldots, e_n\}$. Then there is a unique surjective homomorphism $f\colon F \to M$ such that $f(e_i) = v_i$. The kernel of f is a submodule M_1. Under certain conditions, M_1 is finitely generated (*cf.* Chapter X, §1 on Noetherian rings), and the process can be continued. The systematic study of this process will be carried out in the chapters on resolutions of modules and homology.

Of course, even if M is not finitely generated, one can carry out a similar construction, by using an arbitrary indexing set. Indeed, let $\{v_i\}$ $(i \in I)$ be a family of generators. For each i, let F_i be free with basis consisting of a single element e_i, so $F_i \approx A$. Let F be the direct sum of the modules F_i $(i \in I)$, as in Proposition 3.1. Then we obtain a surjective homomorphism $f\colon F \to M$ such that $f(e_i) = v_i$. Thus every module is a factor module of a free module.

Just as we did for abelian groups in Chapter I, §7, we can also define the **free module** over a ring A **generated by a non-empty set** S. We let $A\langle S\rangle$ be the set of functions $\varphi : S \to A$ such that $\varphi(x) = 0$ for almost all $x \in S$. If $a \in A$ and $x \in S$, we denote by ax the map φ such that $\varphi(x) = a$ and $\varphi(y) = 0$ for $y \neq x$. Then as for abelian groups, given $\varphi \in A\langle S\rangle$ there exist elements $a_i \in A$ and $x_i \in S$ such that

$$\varphi = a_1x_1 + \cdots + a_nx_n.$$

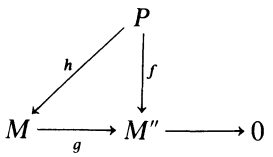
It is immediately verified that the family of functions $\{\delta_x\}$ $(x \in S)$ such that $\delta_x(x) = 1$ and $\delta_x(y) = 0$ for $y \neq x$ form a basis for $A\langle S\rangle$. In other words, the expression of φ as $\sum a_ix_i$ above is unique. This construction can be applied when S is a group or a monoid G, and gives rise to the group algebra as in Chapter II, §5.

Projective modules

There exists another important type of module closely related to free modules, which we now discuss.

Let A be a ring and P a module. The following properties are equivalent, and define what it means for P to be a **projective module**.

P 1. Given a homomorphism $f\colon P \to M''$ and surjective homomorphism $g\colon M \to M''$, there exists a homomorphism $h\colon P \to M$ making the following diagram commutative.



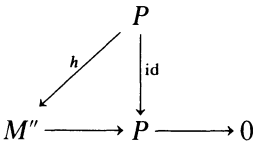
P 2. Every exact sequence $0 \to M' \to M'' \to P \to 0$ splits.

P 3. There exists a module M such that $P \oplus M$ is free, or in words, P is a direct summand of a free module.

P 4. The functor $M \mapsto \mathrm{Hom}_A(P, M)$ is exact.

We prove the equivalence of the four conditions.

Assume **P 1**. Given the exact sequence of **P 2**, we consider the map $f = \mathrm{id}$ in the diagram



Then h gives the desired splitting of the sequence.

Assume **P 2.** Then represent P as a quotient of a free module (cf. Exercise 1) $F \to P \to 0$, and apply **P 2** to this sequence to get the desired splitting, which represents F as a direct sum of P and some module.

Assume **P 3**. Since $\mathrm{Hom}_A(X \oplus Y, M) = \mathrm{Hom}_A(X, M) \oplus \mathrm{Hom}_A(Y, M)$, and since $M \mapsto \mathrm{Hom}_A(F, M)$ is an exact functor if F is free, it follows that $\mathrm{Hom}_A(P, M)$ is exact when P is a direct summand of a free module, which proves **P 4**.

Assume **P 4**. The proof of **P 1** will be left as an exercise.

Examples. It will be proved in the next section that a vector space over a field is always free, i.e. has a basis. Under certain circumstances, it is a theorem that projective modules are free. In §7 we shall prove that a finitely generated projective module over a principal ring is free. In Chapter X, Theorem 4.4 we shall prove that such a module over a local ring is free; in Chapter XVI, Theorem 3.8 we shall prove that a finite flat module over a local ring is free; and in Chapter XXI, Theorem 3.7, we shall prove the Quillen-Suslin theorem that if $A = k[X_1, \ldots, X_n]$ is the polynomial ring over a field k, then every finite projective module over A is free.

Projective modules give rise to the Grothendieck group. Let A be a ring. Isomorphism classes of finite projective modules form a monoid. Indeed, if P is finite projective, let $[P]$ denote its isomorphism class. We define

$$[P] + [Q] = [P \oplus Q].$$

This sum is independent of the choice of representatives P, Q in their class. The conditions defining a monoid are immediately verified. The corresponding Grothendieck group is denoted by $K(A)$.

We can impose a further equivalence relation that P is equivalent to P' if there exist finite free modules F and F' such that $P \oplus F$ is isomorphic to $P' \oplus F'$. Under this equivalence relation we obtain another group denoted by $K_0(A)$. If A is a Dedekind ring (Chapter II, §1 and Exercises 13–19) it can be shown that this group is isomorphic in a natural way with the group of ideal classes $\mathrm{Pic}(A)$ (defined in Chapter II, §1). See Exercises 11, 12, 13. It is also a

problem to determine $K_0(A)$ for as many rings as possible, as explicitly as possible. Algebraic number theory is concerned with $K_0(A)$ when A is the ring of algebraic integers of a number field. The Quillen-Suslin theorem shows if A is the polynomial ring as above, then $K_0(A)$ is trivial.

Of course one can carry out a similar construction with all finite modules. Let $[M]$ denote the isomorphism class of a finite module M. We define the sum to be the direct sum. Then the isomorphism classes of modules over the ring form a monoid, and we can associate to this monoid its Grothendieck group. This construction is applied especially when the ring is commutative. There are many variations on this theme. See for instance the book by Bass, *Algebraic K-theory*, Benjamin, 1968.

There is a variation of the definition of Grothendieck group as follows. Let F be the free abelian group generated by isomorphism classes of finite modules over a ring R, or of modules of bounded cardinality so that we deal with sets. In this free abelian group we let Γ be the subgroup generated by all elements

$$[M] - [M'] - [M'']$$

for which there exists an exact sequence $0 \to M' \to M \to M'' \to 0$. The factor group F/Γ is called the **Grothendieck group** $K(R)$. We shall meet this group again in §8, and in Chapter XX, §3. Note that we may form a similar Grothendieck group with any family of modules such that M is in the family if and only if M' and M'' are in the family. Taking for the family finite projective modules, one sees easily that the two possible definitions of the Grothendieck group coincide in that case.

§5. VECTOR SPACES

A module over a field is called a **vector space**.

Theorem 5.1. *Let V be a vector space over a field K, and assume that $V \neq \{0\}$. Let Γ be a set of generators of V over K and let S be a subset of Γ which is linearly independent. Then there exists a basis $\mathfrak{B}$ of V such that $S \subset \mathfrak{B} \subset \Gamma$.*

Proof. Let $\mathfrak{T}$ be the set whose elements are subsets T of Γ which contain S and are linearly independent. Then $\mathfrak{T}$ is not empty (it contains S), and we contend that $\mathfrak{T}$ is inductively ordered. Indeed, if $\{T_i\}$ is a totally ordered subset

of $\mathfrak{T}$ (by ascending inclusion), then $\bigcup T_i$ is again linearly independent and contains S. By Zorn's lemma, let $\mathfrak{B}$ be a maximal element of $\mathfrak{T}$. Then $\mathfrak{B}$ is linearly independent. Let W be the subspace of V generated by $\mathfrak{B}$. If $W \neq V$, there exists some element $x \in \Gamma$ such that $x \notin W$. Then $\mathfrak{B} \cup \{x\}$ is linearly independent, for given a linear combination

$$\sum_{y \in \mathfrak{B}} a_y y + bx = 0, \qquad a_y, b \in K,$$

we must have $b = 0$, otherwise we get

$$x = -\sum_{y \in \mathfrak{B}} b^{-1} a_y y \in W.$$

By construction, we now see that $a_y = 0$ for all $y \in \mathfrak{B}$, thereby proving that $\mathfrak{B} \cup \{x\}$ is linearly independent, and contradicting the maximality of $\mathfrak{B}$. It follows that $W = V$, and furthermore that $\mathfrak{B}$ is not empty since $V \neq \{0\}$. This proves our theorem.

If V is a vector space $\neq \{0\}$, then in particular, we see that every set of linearly independent elements of V can be extended to a basis, and that a basis may be selected from a given set of generators.

Theorem 5.2. *Let V be a vector space over a field K. Then two bases of V over K have the same cardinality.*

Proof. Let us first assume that there exists a basis of V with a finite number of elements, say $\{v_1, \ldots, v_m\}$, $m \geqq 1$. We shall prove that any other basis must also have m elements. For this it will suffice to prove: If $w_1, \ldots, w_n$ are elements of V which are linearly independent over K, then $n \leqq m$ (for we can then use symmetry). We proceed by induction. There exist elements $c_1, \ldots, c_m$ of K such that

$$w_1 = c_1 v_1 + \cdots + c_m v_m, \tag{1}$$

and some c_i, say c_1, is not equal to 0. Then v_1 lies in the space generated by $w_1, v_2, \ldots, v_m$ over K, and this space must therefore be equal to V itself. Furthermore, $w_1, v_2, \ldots, v_m$ are linearly independent, for suppose $b_1, \ldots, b_m$ are elements of K such that

$$b_1 w_1 + b_2 v_2 + \cdots + b_m v_m = 0.$$

If $b_1 \neq 0$, divide by b_1 and express w_1 as a linear combination of $v_2, \ldots, v_m$. Subtracting from (1) would yield a relation of linear dependence among the v_i, which is impossible. Hence $b_1 = 0$, and again we must have all $b_i = 0$ because the v_i are linearly independent.

Suppose inductively that after a suitable renumbering of the v_i, we have found $w_1, \ldots, w_r$ $(r < n)$ such that

$$\{w_1, \ldots, w_r, v_{r+1}, \ldots, v_m\}$$

is a basis of V. We express w_{r+1} as a linear combination

$$w_{r+1} = c_1 w_1 + \cdots + c_r w_r + c_{r+1} v_{r+1} + \cdots + c_m v_m \tag{2}$$

with $c_i \in K$. The coefficients of the v_i in this relation cannot all be 0; otherwise there would be a linear dependence among the w_j. Say $c_{r+1} \neq 0$. Using an argument similar to that used above, we can replace v_{r+1} by w_{r+1} and still have a basis of V. This means that we can repeat the procedure until $r = n$, and therefore that $n \leqq m$, thereby proving our theorem.

We shall leave the general case of an infinite basis as an exercise to the reader. [*Hint*: Use the fact that a finite number of elements in one basis is contained in the space generated by a finite number of elements in another basis.]

If a vector space V admits one basis with a finite number of elements, say m, then we shall say that V is **finite dimensional** and that m is its **dimension**. In view of Theorem 5.2, we see that m is the number of elements in *any* basis of V. If $V = \{0\}$, then we define its dimension to be 0, and say that V is 0-dimensional. We abbreviate "dimension" by "dim" or "$\dim_K$" if the reference to K is needed for clarity.

When dealing with vector spaces over a field, we use the words **subspace** and **factor space** instead of **submodule** and **factor module**.

Theorem 5.3. *Let V be a vector space over a field K, and let W be a subspace. Then*

$$\dim_K V = \dim_K W + \dim_K V/W.$$

If $f: V \to U$ is a homomorphism of vector spaces over K, then

$$\dim V = \dim \operatorname{Ker} f + \dim \operatorname{Im} f.$$

Proof. The first statement is a special case of the second, taking for f the canonical map. Let $\{u_i\}_{i \in I}$ be a basis of $\operatorname{Im} f$, and let $\{w_j\}_{j \in J}$ be a basis of $\operatorname{Ker} f$. Let $\{v_i\}_{i \in I}$ be a family of elements of V such that $f(v_i) = u_i$ for each $i \in I$. We contend that

$$\{v_i, w_j\}_{i \in I, j \in J}$$

is a basis for V. This will obviously prove our assertion.

Let x be an element of V. Then there exist elements $\{a_i\}_{i \in I}$ of K almost all of which are 0 such that

$$f(x) = \sum_{i \in I} a_i u_i.$$

Hence $f(x - \sum a_i v_i) = f(x) - \sum a_i f(v_i) = 0$. Thus

$$x - \sum a_i v_i$$

is in the kernel of f, and there exist elements $\{b_j\}_{j \in J}$ of K almost all of which are 0 such that

$$x - \sum a_i v_i = \sum b_j w_j.$$

From this we see that $x = \sum a_i v_i + \sum b_j w_j$, and that $\{v_i, w_j\}$ generates V. It remains to be shown that the family $\{v_i, w_j\}$ is linearly independent. Suppose that there exist elements c_i, d_j such that

$$0 = \sum c_i v_i + \sum d_j w_j.$$

Applying f yields

$$0 = \sum c_i f(v_i) = \sum c_i u_i,$$

whence all $c_i = 0$. From this we conclude at once that all $d_j = 0$, and hence that our family $\{v_i, w_j\}$ is a basis for V over K, as was to be shown.

Corollary 5.4. *Let V be a vector space and W a subspace. Then*

$$\dim W \leqq \dim V.$$

If V is finite dimensional and $\dim W = \dim V$ *then $W = V$.*

Proof. Clear.

§6. THE DUAL SPACE AND DUAL MODULE

Let E be a free module over a commutative ring A. We view A as a free module of rank 1 over itself. By the **dual module** $E^\vee$ of E we shall mean the module $\mathrm{Hom}(E, A)$. Its elements will be called **functionals**. Thus a functional on E is an A-linear map $f: E \to A$. If $x \in E$ and $f \in E^\vee$, we sometimes denote $f(x)$ by $\langle x, f \rangle$. Keeping x fixed, we see that the symbol $\langle x, f \rangle$ as a function of $f \in E^\vee$ is A-linear in its second argument, and hence that x induces a linear map on $E^\vee$, which is 0 if and only if $x = 0$. Hence we get an injection $E \to E^{\vee\vee}$ which is not always a surjection.

Let $\{x_i\}_{i \in I}$ be a basis of E. For each $i \in I$ let f_i be the unique functional such that $f_i(x_j) = \delta_{ij}$ (in other words, 1 if $i = j$ and 0 if $i \neq j$). Such a linear map exists by general properties of bases (Theorem 4.1).

Theorem 6.1. *Let E be a finite free module over the commutative ring A, of finite dimension n. Then $E^\vee$ is also free, and* $\dim E^\vee = n$. *If $\{x_1, \ldots, x_n\}$ is a basis for E, and f_i is the functional such that $f_i(x_j) = \delta_{ij}$, then $\{f_1, \ldots, f_n\}$ is a basis for $E^\vee$.*

Proof. Let $f \in E^\vee$ and let $a_i = f(x_i)$ $(i = 1, \ldots, n)$. We have

$$f(c_1x_1 + \cdots + c_nx_n) = c_1f(x_1) + \cdots + c_nf(x_n).$$

Hence $f = a_1f_1 + \cdots + a_nf_n$, and we see that the f_i generate $E^\vee$. Furthermore, they are linearly independent, for if

$$b_1f_1 + \cdots + b_nf_n = 0$$

with $b_i \in K$, then evaluating the left-hand side on x_i yields

$$b_if_i(x_i) = 0,$$

whence $b_i = 0$ for all i. This proves our theorem.

Given a basis $\{x_i\}$ $(i = 1, \ldots, n)$ as in the theorem, we call the basis $\{f_i\}$ the **dual basis**. In terms of these bases, we can express an element A of E with coordinates $(a_1, \ldots, a_n)$, and an element B of $E^\vee$ with coordinates $(b_1, \ldots, b_n)$, such that

$$A = a_1x_1 + \cdots + a_nx_n, \qquad B = b_1f_1 + \cdots + b_nf_n.$$

Then in terms of these coordinates, we see that

$$\langle A, B \rangle = a_1b_1 + \cdots + a_nb_n = A \cdot B$$

is the usual dot product of n-tuples.

Corollary 6.2. *When E is free finite dimensional, then the map $E \to E^{\vee\vee}$ which to each $x \in V$ associates the functional $f \mapsto \langle x, f \rangle$ on $E^\vee$ is an isomorphism of E onto $E^{\vee\vee}$.*

Proof. Note that since $\{f_1, \ldots, f_n\}$ is a basis for $E^\vee$, it follows from the definitions that $\{x_1, \ldots, x_n\}$ is the dual basis in E, so $E = E^{\vee\vee}$.

Theorem 6.3. *Let U, V, W be finite free modules over the commutative ring A, and let*

$$0 \to W \xrightarrow{\lambda} V \xrightarrow{\varphi} U \to 0$$

be an exact sequence of A-homomorphisms. Then the induced sequence

$$0 \to \mathrm{Hom}_A(U, A) \to \mathrm{Hom}_A(V, A) \to \mathrm{Hom}_A(W, A) \to 0$$

i.e.

$$0 \to U^\vee \to V^\vee \to W^\vee \to 0$$

is also exact.

Proof. This is a consequence of **P2**, because a free module is projective.

We now consider properties which have specifically to do with vector spaces, because we are going to take factor spaces. So we assume that we deal with vector spaces over a field K.

Let V, V' be two vector spaces, and suppose given a mapping

$$V \times V' \to K$$

denoted by

$$(x, x') \mapsto \langle x, x' \rangle$$

for $x \in V$ and $x' \in V'$. We call the mapping **bilinear** if for each $x \in V$ the function $x' \mapsto \langle x, x' \rangle$ is linear, and similarly for each $x' \in V'$ the function $x \mapsto \langle x, x' \rangle$ is linear. An element $x \in V$ is said to be **orthogonal** (or **perpendicular**) to a subset S' of V' if $\langle x, x' \rangle = 0$ for all $x' \in S'$. We make a similar definition in the opposite direction. It is clear that the set of $x \in V$ orthogonal to S' is a subspace of V.

We define the **kernel** of the bilinear map on the left to be the subspace of V which is orthogonal to V', and similarly for the kernel on the right.

Given a bilinear map as above,

$$V \times V' \to K,$$

let W' be its kernel on the right and let W be its kernel on the left. Let x' be an element of V'. Then x' gives rise to a functional on V, by the rule $x \mapsto \langle x, x' \rangle$, and this functional obviously depends only on the coset of x' modulo W'; in other words, if $x'_1 \equiv x'_2 \pmod{W'}$, then the functionals $x \mapsto \langle x, x'_1 \rangle$ and $x \mapsto \langle x, x'_2 \rangle$ are equal. Hence we get a homomorphism

$$V' \to V^\vee$$

whose kernel is precisely W' by definition, whence an injective homomorphism

$$0 \to V'/W' \to V^\vee.$$

Since all the functionals arising from elements of V' vanish on W, we can view them as functionals on V/W, i.e. as elements of $(V/W)^\vee$. So we actually get an injective homomorphism

$$0 \to V'/W' \to (V/W)^\vee.$$

One could give a name to the homomorphism

$$g : V' \to V^\vee$$

such that

$$\langle x, x' \rangle = \langle x, g(x') \rangle$$

for all $x \in V$ and $x' \in V'$. However, it will usually be possible to describe it by an arrow and call it the induced map, or the natural map. Giving a name to it would tend to make the terminology heavier than necessary.

Theorem 6.4. *Let $V \times V' \to K$ be a bilinear map, let W, W' be its kernels on the left and right respectively, and assume that V'/W' is finite dimensional. Then the induced homomorphism $V'/W' \to (V/W)^\vee$ is an isomorphism.*

Proof. By symmetry, we have an induced homomorphism

$$V/W \to (V'/W')^\vee$$

which is injective. Since

$$\dim(V'/W')^\vee = \dim V'/W'$$

it follows that V/W is finite dimensional. From the above injective homomorphism and the other, namely

$$0 \to V'/W' \to (V/W)^\vee,$$

we get the inequalities

$$\dim V/W \leqq \dim V'/W'$$

and

$$\dim V'/W' \leqq \dim V/W,$$

whence an equality of dimensions. Hence our homomorphisms are surjective and inverse to each other, thereby proving the theorem.

Remark 1. Theorem 6.4 is the analogue for vector spaces of the duality Theorem 9.2 of Chapter I.

Remark 2. Let A be a commutative ring and let E be an A-module. Then we may form two types of dual:

$E^\wedge = \operatorname{Hom}(E, \mathbf{Q}/\mathbf{Z})$, viewing E as an abelian group;

$E^\vee = \operatorname{Hom}_A(E, A)$, viewing E as an A-module.

Both are called **dual**, and they usually are applied in different contexts. For instance, $E^\vee$ will be considered in Chapter XIII, while $E^\wedge$ will be considered in the theory of injective modules, Chapter XX, §4. For an example of dual module $E^\vee$ see Exercise 11. If by any chance the two duals arise together and there is need to distinguish between them, then we may call $E^\wedge$ the **Pontrjagin dual**.

Indeed, in the theory of topological groups G, the group of continuous homomorphisms of G into $\mathbf{R}/\mathbf{Z}$ is the classical Pontrjagin dual, and is classically denoted by $G^\wedge$, so I find the preservation of that terminology appropriate.

Instead of $\mathbf{R}/\mathbf{Z}$ one may take other natural groups isomorphic to $\mathbf{R}/\mathbf{Z}$. The most common such group is the group of complex numbers of absolute value 1, which we denote by $\mathbf{S}^1$. The isomorphism with $\mathbf{R}/\mathbf{Z}$ is given by the map

$$x \mapsto e^{2\pi i x}.$$

Remark 3. A bilinear map $V \times V \to K$ for which $V' = V$ is called a **bilinear form**. We say that the form is **non-singular** if the corresponding maps

$$V' \to V^\vee \quad \text{and} \quad V \to (V')^\vee$$

are isomorphisms. Bilinear maps and bilinear forms will be studied at greater length in Chapter XV. See also Exercise 33 of Chapter XIII for a nice example.

§7. MODULES OVER PRINCIPAL RINGS

Throughout this section, we assume that R is a principal entire ring. All modules are over R, and homomorphisms are R-homomorphisms, unless otherwise specified.

The theorems will generalize those proved in Chapter I for abelian groups. We shall also point out how the proofs of Chapter I can be adjusted with substitutions of terminology so as to yield proofs in the present case.

Let F be a free module over R, with a basis $\{x_i\}_{i \in I}$. Then the cardinality of I is uniquely determined, and is called the **dimension** of F. We recall that this is proved, say by taking a prime element p in R, and observing that F/pF is a vector space over the field R/pR, whose dimension is precisely the cardinality of I. We may therefore speak of the dimension of a free module over R.

Theorem 7.1. *Let F be a free module, and M a submodule. Then M is free, and its dimension is less than or equal to the dimension of F.*

Proof. For simplicity, we give the proof when F has a finite basis $\{x_i\}$, $i = 1, \ldots, n$. Let M_r be the intersection of M with $(x_1, \ldots, x_r)$, the module generated by $x_1, \ldots, x_r$. Then $M_1 = M \cap (x_1)$ is a submodule of (x_1), and is therefore of type $(a_1 x_1)$ with some $a_1 \in R$. Hence M_1 is either 0 or free, of dimension 1. Assume inductively that M_r is free of dimension $\leqq r$. Let $\mathfrak{a}$ be the set consisting of all elements $a \in R$ such that there exists an element $x \in M$ which can be written

$$x = b_1 x_1 + \cdots + b_r x_r + a x_{r+1}$$

with $b_i \in R$. Then $\mathfrak{a}$ is obviously an ideal, and is principal, generated say by an element a_{r+1}. If $a_{r+1} = 0$, then $M_{r+1} = M_r$ and we are done with the inductive step. If $a_{r+1} \neq 0$, let $w \in M_{r+1}$ be such that the coefficient of w with respect to x_{r+1} is a_{r+1}. If $x \in M_{r+1}$ then the coefficient of x with respect to x_{r+1} is divisible by a_{r+1}, and hence there exists $c \in R$ such that $x - cw$ lies in M_r. Hence

$$M_{r+1} = M_r + (w).$$

On the other hand, it is clear that $M_r \cap (w)$ is 0, and hence that this sum is direct, thereby proving our theorem. (For the infinite case, see Appendix 2, §2.)

Corollary 7.2. *Let E be a finitely generated module and E' a submodule. Then E' is finitely generated.*

Proof. We can represent E as a factor module of a free module F with a finite number of generators: If $v_1, \ldots, v_n$ are generators of E, we take a free module F with basis $\{x_1, \ldots, x_n\}$ and map x_i on v_i. The inverse image of E' in F is a submodule, which is free, and finitely generated, by the theorem. Hence E' is finitely generated. The assertion also follows using simple properties of Noetherian rings and modules.

If one wants to translate the proofs of Chapter I, then one makes the following definitions. A free 1-dimensional module over R is called **infinite cyclic**. An infinite cyclic module is isomorphic to R, viewed as module over itself. Thus every non-zero submodule of an infinite cyclic module is infinite cyclic. The proof given in Chapter I for the analogue of Theorem 7.1 applies without further change.

Let E be a module. We say that E is a **torsion** module if given $x \in E$, there exists $a \in R$, $a \neq 0$, such that $ax = 0$. The generalization of **finite abelian group** is **finitely generated torsion module**. An element x of E is called a **torsion element** if there exists $a \in R$, $a \neq 0$, such that $ax = 0$.

Let E be a module. We denote by E_{tor} the submodule consisting of all torsion elements of E, and call it the **torsion submodule** of E. If $E_{\mathrm{tor}} = 0$, we say that E is **torsion free**.

Theorem 7.3. *Let E be finitely generated. Then E/E_{tor} is free. There exists a free submodule F of E such that E is a direct sum*

$$E = E_{\mathrm{tor}} \oplus F.$$

The dimension of such a submodule F is uniquely determined.

Proof. We first prove that E/E_{tor} is torsion free. If $x \in E$, let $\bar{x}$ denote its residue class mod E_{tor}. Let $b \in R$, $b \neq 0$ be such that $b\bar{x} = 0$. Then $bx \in E_{\mathrm{tor}}$, and hence there exists $c \in R$, $c \neq 0$, such that $cbx = 0$. Hence $x \in E_{\mathrm{tor}}$ and $\bar{x} = 0$, thereby proving that E/E_{tor} is torsion free. It is also finitely generated.

Assume now that M is a torsion free module which is finitely generated. Let $\{v_1, \ldots, v_n\}$ be a maximal set of elements of M among a given finite set of generators $\{y_1, \ldots, y_m\}$ such that $\{v_1, \ldots, v_n\}$ is linearly independent. If y is one of the generators, there exist elements $a, b_1, \ldots, b_n \in R$ not all 0, such that

$$ay + b_1v_1 + \cdots + b_nv_n = 0.$$

Then $a \neq 0$ (otherwise we contradict the linear independence of $v_1, \ldots, v_n$). Hence ay lies in $(v_1, \ldots, v_n)$. Thus for each $j = 1, \ldots, m$ we can find $a_j \in R$, $a_j \neq 0$, such that a_jy_j lies in $(v_1, \ldots, v_n)$. Let $a = a_1 \cdots a_m$ be the product. Then aM is contained in $(v_1, \ldots, v_n)$, and $a \neq 0$. The map

$$x \mapsto ax$$

is an injective homomorphism, whose image is contained in a free module. This image is isomorphic to M, and we conclude from Theorem 7.1 that M is free, as desired.

To get the submodule F we need a lemma.

Lemma 7.4. *Let E, E' be modules, and assume that E' is free. Let $f : E \to E'$ be a surjective homomorphism. Then there exists a free submodule F of E such that the restriction of f to F induces an isomorphism of F with E', and such that $E = F \oplus \operatorname{Ker} f$.*

Proof. Let $\{x_i'\}_{i \in I}$ be a basis of E'. For each i, let x_i be an element of E such that $f(x_i) = x_i'$. Let F be the submodule of E generated by all the elements x_i, $i \in I$. Then one sees at once that the family of elements $\{x_i\}_{i \in I}$ is linearly independent, and therefore that F is free. Given $x \in E$, there exist elements $a_i \in R$ such that

$$f(x) = \sum a_i x_i'.$$

Then $x - \sum a_i x_i$ lies in the kernel of f, and therefore $E = \operatorname{Ker} f + F$. It is clear that $\operatorname{Ker} f \cap F = 0$, and hence that the sum is direct, thereby proving the lemma.

We apply the lemma to the homomorphism $E \to E/E_{\text{tor}}$ in Theorem 7.3 to get our decomposition $E = E_{\text{tor}} \oplus F$. The dimension of F is uniquely determined, because F is isomorphic to E/E_{tor} for any decomposition of E into a direct sum as stated in the theorem.

The dimension of the free module F in Theorem 7.3 is called the **rank** of E.

In order to get the structure theorem for finitely generated modules over R, one can proceed exactly as for abelian groups. We shall describe the dictionary which allows us to transport the proofs essentially without change.

Let E be a module over R. Let $x \in E$. The map $a \mapsto ax$ is a homomorphism of R onto the submodule generated by x, and the kernel is an ideal, which is principal, generated by an element $m \in R$. We say that m is a **period** of x. We

note that m is determined up to multiplication by a unit (if $m \neq 0$). An element $c \in R, c \neq 0$, is said to be an **exponent** for E (resp. for x) if $cE = 0$ (resp. $cx = 0$).

Let p be a prime element. We denote by $E(p)$ the submodule of E consisting of all elements x having an exponent which is a power p^r ($r \geqq 1$). A p-submodule of E is a submodule contained in $E(p)$.

We select once and for all a system of representatives for the prime elements of R (modulo units). For instance, if R is a polynomial ring in one variable over a field, we take as representatives the irreducible polynomials with leading coefficient 1.

Let $m \in R, m \neq 0$. We denote by E_m the kernel of the map $x \mapsto mx$. It consists of all elements of E having exponent m.

A module E is said to be **cyclic** if it is isomorphic to $R/(a)$ for some element $a \in R$. Without loss of generality if $a \neq 0$, one may assume that a is a product of primes in our system of representatives, and then we could say that a is the order of the module.

Let $r_1, \ldots, r_s$ be integers $\geqq 1$. A p-module E is said to be of **type**

$$(p^{r_1}, \ldots, p^{r_s})$$

if it is isomorphic to the product of cyclic modules $R/(p^{r_i})$ ($i = 1, \ldots, s$). If p is fixed, then one could say that the module is of type $(r_1, \ldots, r_s)$ (relative to p).

All the proofs of Chapter I, §8 now go over without change. Whenever we argue on the size of a positive integer m, we have a similar argument on the number of prime factors appearing in its prime factorization. If we deal with a prime power p^r, we can view the order as being determined by r. The reader can now check that the proofs of Chapter I, §8 are applicable.

However, we shall develop the theory once again without assuming any knowledge of Chapter I, §8. Thus our treatment is self-contained.

Theorem 7.5. *Let E be a finitely generated torsion module $\neq 0$. Then E is the direct sum*

$$E = \bigoplus_p E(p),$$

taken over all primes p such that $E(p) \neq 0$. Each $E(p)$ can be written as a direct sum

$$E(p) = R/(p^{v_1}) \oplus \cdots \oplus R/(p^{v_s})$$

with $1 \leqq v_1 \leqq \cdots \leqq v_s$. The sequence $v_1, \ldots, v_s$ is uniquely determined.

Proof. Let a be an exponent for E, and suppose that $a = bc$ with $(b, c) = (1)$. Let $x, y \in R$ be such that

$$1 = xb + yc.$$

We contend that $E = E_b \oplus E_c$. Our first assertion then follows by induction, expressing a as a product of prime powers. Let $v \in E$. Then

$$v = xbv + ycv.$$

Then $xbv \in E_c$ because $cxbv = xav = 0$. Similarly, $ycv \in E_b$. Finally $E_b \cap E_c = 0$, as one sees immediately. Hence E is the direct sum of E_b and E_c.

We must now prove that $E(p)$ is a direct sum as stated. If $y_1, \ldots, y_m$ are elements of a module, we shall say that they are **independent** if whenever we have a relation

$$a_1 y_1 + \cdots + a_m y_m = 0$$

with $a_i \in R$, then we must have $a_i y_i = 0$ for all i. (Observe that **independent** does not mean **linearly independent**.) We see at once that $y_1, \ldots, y_m$ are independent if and only if the module $(y_1, \ldots, y_m)$ has the direct sum decomposition

$$(y_1, \ldots, y_m) = (y_1) \oplus \cdots \oplus (y_m)$$

in terms of the cyclic modules (y_i), $i = 1, \ldots, m$.

We now have an analogue of Lemma 7.4 for modules having a prime power exponent.

Lemma 7.6. *Let E be a torsion module of exponent p^r ($r \geqq 1$) for some prime element p. Let $x_1 \in E$ be an element of period p^r. Let $\bar{E} = E/(x_1)$. Let $\bar{y}_1, \ldots, \bar{y}_m$ be independent elements of $\bar{E}$. Then for each i there exists a representative $y_i \in E$ of $\bar{y}_i$, such that the period of y_i is the same as the period of $\bar{y}_i$. The elements $x_1, y_1, \ldots, y_m$ are independent.*

Proof. Let $\bar{y} \in \bar{E}$ have period p^n for some $n \geqq 1$. Let y be a representative of $\bar{y}$ in E. Then $p^n y \in (x_1)$, and hence

$$p^n y = p^s c x_1, \qquad c \in R, p \nmid c,$$

for some $s \leqq r$. If $s = r$, we see that y has the same period as $\bar{y}$. If $s < r$, then $p^s c x_1$ has period p^{r-s}, and hence y has period p^{n+r-s}. We must have

$$n + r - s \leqq r,$$

because p^r is an exponent for E. Thus we obtain $n \leqq s$, and we see that

$$y - p^{s-n} c x_1$$

is a representative for $\bar{y}$, whose period is p^n.

Let y_i be a representative for $\bar{y}_i$ having the same period. We prove that $x_1, y_1, \ldots, y_m$ are independent. Suppose that $a, a_1, \ldots, a_m \in R$ are elements such that

$$a x_1 + a_1 y_1 + \cdots + a_m y_m = 0.$$

Then

$$a_1\bar{y}_1 + \cdots + a_m\bar{y}_m = 0.$$

By hypothesis, we must have $a_i\bar{y}_i = 0$ for each i. If p^{r_i} is the period of $\bar{y}_i$, then p^{r_i} divides a_i. We then conclude that $a_i y_i = 0$ for each i, and hence finally that $ax_1 = 0$, thereby proving the desired independence.

To get the direct sum decomposition of $E(p)$, we first note that $E(p)$ is finitely generated. We may assume without loss of generality that $E = E(p)$. Let x_1 be an element of E whose period p^{r_1} is such that r_1 is maximal. Let $\bar{E} = E/(x_1)$. We contend that $\dim \bar{E}_p$ as vector space over R/pR is strictly less than $\dim E_p$. Indeed, if $\bar{y}_1, \ldots, \bar{y}_m$ are linearly independent elements of $\bar{E}_p$ over R/pR, then Lemma 7.6 implies that $\dim E_p \geqq m + 1$ because we can always find an element of (x_1) having period p, independent of $y_1, \ldots, y_m$. Hence $\dim \bar{E}_p < \dim E_p$. We can prove the direct sum decomposition by induction. If $E \neq 0$, there exist elements $\bar{x}_2, \ldots, \bar{x}_s$ having periods $p^{r_2}, \ldots, p^{r_s}$ respectively, such that $r_2 \geqq \cdots \geqq r_s$. By Lemma 7.6, there exist representatives $x_2, \ldots, x_r$ in E such that x_i has period p^{r_i} and $x_1, \ldots, x_r$ are independent. Since p^{r_1} is such that r_1 is maximal, we have $r_1 \geqq r_2$, and our decomposition is achieved.

The uniqueness will be a consequence of a more general uniqueness theorem, which we state next.

Theorem 7.7. *Let E be a finitely generated torsion module, $E \neq 0$. Then E is isomorphic to a direct sum of non-zero factors*

$$R/(q_1) \oplus \cdots \oplus R/(q_r),$$

where $q_1, \ldots, q_r$ are non-zero non-units of R, and $q_1 \mid q_2 \mid \cdots \mid q_r$. The sequence of ideals $(q_1), \ldots, (q_r)$ is uniquely determined by the above conditions.

Proof. Using Theorem 7.5, decompose E into a direct sum of p-submodules, say $E(p_1) \oplus \cdots \oplus E(p_l)$, and then decompose each $E(p_i)$ into a direct sum of cyclic submodules of periods $p_i^{r_{ij}}$. We visualize these symbolically as described by the following diagram:

$$\begin{array}{lcccc} E(p_1): & r_{11} & \leqq & r_{12} & \leqq \cdots \\ E(p_2): & r_{21} & \leqq & r_{22} & \leqq \cdots \\ \vdots & \vdots & & \vdots & \vdots \\ E(p_l): & r_{l1} & \leqq & r_{l2} & \leqq \cdots \end{array}$$

A horizontal row describes the type of the module with respect to the prime at the left. The exponents r_{ij} are arranged in increasing order for each fixed $i = 1, \ldots, l$. We let $q_1, \ldots, q_r$ correspond to the columns of the matrix of exponents, in other words

$$q_1 = p_1^{r_{11}} p_2^{r_{21}} \cdots p_l^{r_{l1}},$$

$$q_2 = p_1^{r_{12}} p_2^{r_{22}} \cdots p_l^{r_{l2}},$$

$$\cdots$$

The direct sum of the cyclic modules represented by the first column is then isomorphic to $R/(q_1)$, because, as with abelian groups, the direct sum of cyclic modules whose periods are relatively prime is also cyclic. We have a similar remark for each column, and we observe that our proof actually orders the q_j by increasing divisibility, as was to be shown.

Now for uniqueness. Let p be any prime, and suppose that $E = R/(pb)$ for some $b \in R$, $b \neq 0$. Then E_p is the submodule $bR/(pb)$, as follows at once from unique factorization in R. But the kernel of the composite map

$$R \to bR \to bR/(pb)$$

is precisely (p). Thus we have an isomorphism

$$R/(p) \approx bR/(pb).$$

Let now E be expressed as in the theorem, as a direct sum of r terms. An element

$$v = v_1 \oplus \cdots \oplus v_r, \qquad v_i \in R/(q_i)$$

is in E_p if and only if $pv_i = 0$ for all i. Hence E_p is the direct sum of the kernel of multiplication by p in each term. But E_p is a vector space over $R/(p)$, and its dimension is therefore equal to the number of terms $R/(q_i)$ such that p divides q_i.

Suppose that p is a prime dividing q_1, and hence q_i for each $i = 1, \ldots, r$. Let E have a direct sum decomposition into d terms satisfying the conditions of the theorem, say

$$E = R/(q'_1) \oplus \cdots \oplus R/(q'_s).$$

Then p must divide at least r of the elements q'_j, whence $r \leqq s$. By symmetry, $r = s$, and p divides q'_j for all j.

Consider the module pE. By a preceding remark, if we write $q_i = pb_i$, then

$$pE \approx R/(b_1) \oplus \cdots \oplus R/(b_r),$$

and $b_1 | \cdots | b_r$. Some of the b_i may be units, but those which are not units determine their principal ideal uniquely, by induction. Hence if

$$(b_1) = \cdots = (b_j) = 1$$

but $(b_{j+1}) \neq (1)$, then the sequence of ideals

$$(b_{j+1}), \ldots, (b_r)$$

is uniquely determined. This proves our uniqueness statement, and concludes the proof of Theorem 7.7.

The ideals $(q_1), \ldots, (q_r)$ are called the **invariants** of E.

For one of the main applications of Theorem 7.7 to linear algebra, see Chapter XV, §2.

The next theorem is included for completeness. It is called the **elementary divisors** theorem.

Theorem 7.8. *Let F be a free module over R, and let M be a finitely generated submodule $\neq 0$. Then there exists a basis $\mathfrak{B}$ of F, elements $e_1, \ldots, e_m$ in this basis, and non-zero elements $a_1, \ldots, a_m \in R$ such that:*

(i) *The elements $a_1e_1, \ldots, a_me_m$ form a basis of M over R.*

(ii) *We have $a_i | a_{i+1}$ for $i = 1, \ldots, m-1$.*

The sequence of ideals $(a_1), \ldots, (a_m)$ is uniquely determined by the preceding conditions.

Proof. Write a finite set of generators for M as linear combination of a finite number of elements in a basis for F. These elements generate a free submodule of finite rank, and thus it suffices to prove the theorem when F has finite rank, which we now assume. We let $n = \operatorname{rank}(F)$.

The uniqueness is a corollary of Theorem 7.7. Suppose we have a basis as in the theorem. Say $a_1, \ldots, a_s$ are units, and so can be taken to be $= 1$, and $a_{s+j} = q_j$ with $q_1 | q_2 | \ldots | q_r$ non-units. Observe that $F/M = \bar{F}$ is a finitely generated module over R, having the direct sum expression

$$F/M = \bar{F} \approx \bigoplus_{j=1}^{r} (R/q_jR)\bar{e}_j \oplus \text{free module of rank } n - (r+s)$$

where a bar denotes the class of an element of F mod M. Thus the direct sum over $j = 1, \ldots, r$ is the torsion submodule of $\bar{F}$, whence the elements $q_1, \ldots, q_r$ are uniquely determined by Theorem 7.7. We have $r + s = m$, so the rank of F/M is $n - m$, which determines m uniquely. Then $s = m - r$ is uniquely determined as the number of units among $a_1, \ldots, a_m$. This proves the uniqueness part of the theorem. Next we prove existence.

Let λ be a functional on F, in other words, an element of $\operatorname{Hom}_R(F, R)$. We let $J_\lambda = \lambda(M)$. Then J_λ is an ideal of R. Select λ_1 such that $\lambda_1(M)$ is maximal in the set of ideals $\{J_\lambda\}$, that is to say, there is no properly larger ideal in the set $\{J_\lambda\}$.

Let $\lambda_1(M) = (a_1)$. Then $a_1 \neq 0$, because there exists a non-zero element of M, and expressing this element in terms of some basis for F over R, with some non-zero coordinate, we take the projection on this coordinate to get a functional whose value on M is not 0. Let $x_1 \in M$ be such that $\lambda_1(x_1) = a_1$. For any functional g we must have $g(x_1) \in (a_1)$ [immediate from the maximality of

$\lambda_1(M)$]. Writing x_1 in terms of any basis of F, we see that its coefficients must all be divisible by a_1. (If some coefficient is not divisible by a_1, project on this coefficient to get an impossible functional.) Therefore we can write $x_1 = a_1 e_1$ with some element $e_1 \in F$.

Next we prove that F is a direct sum

$$F = Re_1 \oplus \operatorname{Ker} \lambda_1.$$

Since $\lambda_1(e_1) = 1$, it is clear that $Re_1 \cap \operatorname{Ker} \lambda_1 = 0$. Furthermore, given $x \in F$ we note that $x - \lambda_1(x)e_1$ is in the kernel of λ_1. Hence F is the sum of the indicated submodules, and therefore the direct sum.

We note that $\operatorname{Ker} \lambda_1$ is free, being a submodule of a free module (Theorem 7.1). We let

$$F_1 = \operatorname{Ker} \lambda_1 \quad \text{and} \quad M_1 = M \cap \operatorname{Ker} \lambda_1.$$

We see at once that $M = Rx_1 \oplus M_1$.

Thus M_1 is a submodule of F_1 and its dimension is one less than the dimension of M. From the maximality condition on $\lambda_1(M)$, it follows at once that for any functional λ on F_1, the image $\lambda(M)$ will be contained in $\lambda_1(M)$ (because otherwise, a suitable linear combination of functionals would yield an ideal larger than (a_1)). We can therefore complete the existence proof by induction.

In Theorem 7.8, we call the ideals $(a_1), \ldots, (a_m)$ the **invariants** of M in F. For another characterization of these invariants, see Chapter XIII, Proposition 4.20.

Example. First, see examples of situations similar to those of Theorem 7.8 in Exercises 5, 7, and 8, and for Dedekind rings in Exercise 13.

Example. Another way to obtain a module M as in Theorem 7.8 is as a module of relations. Let W be a finitely generated module over R, with generators $w_1, \ldots, w_n$. By a **relation** among $\{w_1, \ldots, w_n\}$ we mean an element $(a_1, \ldots, a_n) \in R^n$ such that $\sum a_i w_i = 0$. The set of such relations is a submodule of R^n, to which Theorem 7.8 may be applied.

It is also possible to formulate a proof of Theorem 7.8 by considering M as a submodule of R^n, and applying the method of row and column operations to get a desired basis. In this context, we make some further comments which may serve to illustrate Theorem 7.8. We assume that the reader is acquainted with matrices over a ring. By **row operations** we mean: interchanging two rows; adding a multiple of one row to another; multiplying a row by a unit in the ring. We define **column operations** similarly. These row and column operations correspond to multiplication with the so-called elementary matrices in the ring.

Theorem 7.9. *Assume that the elementary matrices in R generate $GL_n(R)$. Let (x_{ij}) be a non-zero matrix with components in R. Then with a finite number of row and column operations, it is possible to bring the matrix to the form*

$$\begin{pmatrix} a_1 & 0 & \cdots & \cdot & \cdot & \cdots & 0 \\ 0 & a_2 & \cdots & \cdot & \cdot & \cdots & 0 \\ \vdots & & \ddots & & & & \vdots \\ 0 & \cdot & \cdots & a_m & \cdot & \cdots & 0 \\ 0 & \cdot & \cdots & \cdot & 0 & \cdots & 0 \\ \vdots & & & & & & \vdots \\ 0 & \cdot & \cdots & \cdot & \cdot & \cdots & 0 \end{pmatrix}.$$

with $a_1 \cdots a_m \neq 0$ *and* $a_1 \mid a_2 \mid \cdots \mid a_m$.

We leave the proof for the reader. Either Theorem 7.9 can be viewed as equivalent to Theorem 7.8, or a direct proof may be given. In any case, Theorem 7.9 can be used in the following context. Consider a system of linear equations

$$c_{11}x_1 + \cdots + c_{1n}x_n = 0$$
$$\cdots$$
$$c_{r1}x_1 + \cdots + c_{rn}x_n = 0.$$

with coefficients in R. Let F be the submodule of R^n generated by the vectors $X = (x_1, \ldots, x_n)$ which are solutions of this system. By Theorem 7.1, we know that F is free of dimension $\leqq n$. Theorem 7.9 can be viewed as providing a normalized basis for F in line with Theorem 7.8.

Further example. As pointed out by Paul Cohen, the row and column method can be applied to modules over a power series ring $\mathfrak{o}[[X]]$, where $\mathfrak{o}$ is a complete discrete valuation ring. Cf. Theorem 3.1 of Chapter 5 in my *Cyclotomic Fields* I *and* II (Springer Verlag, 1990). For instance, one could pick $\mathfrak{o}$ itself to be a power series ring $k[[T]]$ in one variable over a field k, but in the theory of cyclotomic fields in the above reference, $\mathfrak{o}$ is taken to be the ring of p-adic integers. On the other hand, George Bergman has drawn my attention to P. M. Cohn's "On the structure of GL_2 of a ring," *IHES Publ. Math.* No. 30 (1966), giving examples of principal rings where one cannot use row and column operations in Theorem 7.9.

§8. EULER-POINCARÉ MAPS

The present section may be viewed as providing an example and application of the Jordan-Hölder theorem for modules. But as pointed out in the examples and references below, it also provides an introduction for further theories.

Again let A be a ring. We continue to consider A-modules. Let Γ be an abelian group, written additively. Let φ be a rule which to certain modules associates an element of Γ, subject to the following condition:

If $0 \to M' \to M \to M'' \to 0$ *is exact, then* $\varphi(M)$ *is defined if and only if* $\varphi(M')$ *and* $\varphi(M'')$ *are defined, and in that case, we have*

$$\varphi(M) = \varphi(M') + \varphi(M'').$$

Furthermore $\varphi(0)$ *is defined and equal to* 0.

Such a rule φ will be called an **Euler-Poincaré mapping** on the category of A-modules. If M' is isomorphic to M, then from the exact sequence

$$0 \to M' \to M \to 0 \to 0$$

we conclude that $\varphi(M')$ is defined if $\varphi(M)$ is defined, and that $\varphi(M') = \varphi(M)$. Thus if $\varphi(M)$ is defined for a module M, φ is defined on every submodule and factor module of M. In particular, if we have an exact sequence of modules

$$M' \to M \to M''$$

and if $\varphi(M')$ and $\varphi(M'')$ are defined, then so is $\varphi(M)$, as one sees at once by considering the kernel and image of our two maps, and using the definition.

Examples. We could let $A = \mathbf{Z}$, and let φ be defined for all finite abelian groups, and be equal to the order of the group. The value of φ is in the multiplicative group of positive rational numbers.

As another example, we consider the category of vector spaces over a field k. We let φ be defined for finite dimensional spaces, and be equal to the dimension. The values of φ are then in the additive group of integers.

In Chapter XV we shall see that the characteristic polynomial may be considered as an Euler-Poincaré map.

Observe that the natural map of a finite module into its image in the Grothendieck group defined at the end of §4 is a universal Euler-Poincaré mapping. We shall develop a more extensive theory of this mapping in Chapter XX, §3.

If M is a module (over a ring A), then a sequence of submodules

$$M = M_1 \supset M_2 \supset \cdots \supset M_r = 0$$

is also called a **finite filtration**, and we call r the **length** of the filtration. A module M is said to be **simple** if it does not contain any submodule other than 0 and M itself, and if $M \neq 0$. A filtration is said to be **simple** if each M_i/M_{i+1} is simple. *The Jordan-Hölder theorem asserts that two simple filtrations of a module are equivalent.*

A module M is said to be of **finite length** if it is 0 or if it admits a simple (finite) filtration. By the Jordan-Hölder theorem for modules (proved the same way as for groups), the length of such a simple filtration is uniquely determined, and is called the **length of the module**. In the language of Euler characteristics, the Jordan-Hölder theorem can be reformulated as follows:

Theorem 8.1. *Let φ be a rule which to each simple module associates an element of a commutative group Γ, and such that if $M \approx M'$ then*

$$\varphi(M) = \varphi(M').$$

Then φ has a unique extension to an Euler-Poincaré mapping defined on all modules of finite length.

Proof. Given a simple filtration

$$M = M_1 \supset M_2 \supset \cdots \supset M_r = 0$$

we define

$$\varphi(M) = \sum_{i=1}^{r-1} \varphi(M_i/M_{i+1}).$$

The Jordan-Hölder theorem shows immediately that this is well-defined, and that this extension of φ is an Euler-Poincaré map.

In particular, we see that the length function is the Euler-Poincaré map taking its values in the additive group of integers, and having the value 1 for any simple module.

§9. THE SNAKE LEMMA

This section gives a very general lemma, which will be used many times, so we extract it here. The reader may skip it until it is encountered, but already we give some exercises which show how it is applied: the five lemma in Exercise 15 and also Exercise 26. Other substantial applications in this book will occur in Chapter XVI, §3 in connection with the tensor product, and in Chapter XX in connection with complexes, resolutions, and derived functors.

We begin with routine comments. Consider a commutative diagram of homomorphisms of modules.

$$\begin{array}{ccc} M' & \xrightarrow{f} & M \\ {\scriptstyle d'}\downarrow & & \downarrow{\scriptstyle d} \\ N' & \xrightarrow[h]{} & N \end{array}$$

Then f induces a homomorphism

$$\operatorname{Ker} d' \to \operatorname{Ker} d.$$

Indeed, suppose $d'x' = 0$. Then $df(x') = 0$ because $df(x') = hd'(x') = 0$.

Similarly, h induces a homomorphism

$$\text{Coker } d' \to \text{Coker } d$$

in a natural way as follows. Let $y' \in N'$ represent an element of $N'/d'M'$. Then $hy' \bmod dM$ does not depend on the choice of y' representing the given element, because if $y'' = y' + d'x'$, then

$$hy'' = hy' + hd'x' = hy' + dfx' \equiv hy' \bmod dM.$$

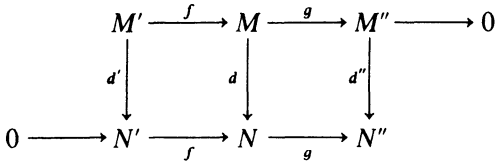
Thus we get a map

$$h_*: N'/d'M' = \text{Coker } d' \to N/dM = \text{Coker } d,$$

which is immediately verified to be a homomorphism.

In practice, given a commutative diagram as above, one sometimes writes f instead of h, so one writes f for the horizontal maps both above and below the diagram. This simplifies the notation, and is not so incorrect: we may view M', N' as the two components of a direct sum, and similarly for M, N. Then f is merely a homomorphism defined on the direct sum $M' \oplus N'$ into $M \oplus N$.

The snake lemma concerns a commutative and exact diagram called a **snake diagram**:



Let $z'' \in \text{Ker } d''$. We can construct elements of N' as follows. Since g is surjective, there exists an element $z \in M$ such that $gz = z''$. We now move vertically down by d, and take dz. The commutativity $d''g = gd$ shows that $gdz = 0$ whence dz is in the kernel of g in N. By exactness, there exists an element $z' \in N'$ such that $fz' = dz$. In brief, we write

$$z' = f^{-1} \circ d \circ g^{-1} z''.$$

Of course, z' is not well defined because of the choices made when taking inverse images. However, the snake lemma will state exactly what goes on.

Lemma 9.1. (Snake Lemma). *Given a snake diagram as above, the map*

$$\delta : \text{Ker } d'' \to \text{Coker } d'$$

induced by $\delta z'' = f^{-1} \circ d \circ g^{-1} z''$ *is well defined, and we have an exact sequence*

$$\text{Ker } d' \to \text{Ker } d \to \text{Ker } d'' \xrightarrow{\delta} \text{Coker } d' \to \text{Coker } d \to \text{Coker } d''$$

where the maps besides δ *are the natural ones.*

Proof. It is a routine verification that the class of $z' \bmod \operatorname{Im} d'$ is independent of the choices made when taking inverse images, whence defining the map δ. The proof of the exactness of the sequence is then routine, and consists in chasing around diagrams. It should be carried out in full detail by the reader who wishes to acquire a feeling for this type of triviality. As an example, we shall prove that

$$\operatorname{Ker} \delta \subset \operatorname{Im} g_*$$

where g_* is the induced map on kernels. Suppose the image of z'' is 0 in Coker d'. By definition, there exists $u' \in M'$ such that $z' = d'u'$. Then

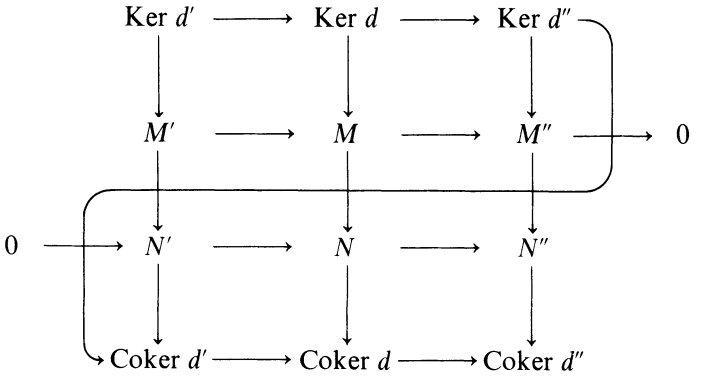
$$dz = fz' = fd'u' = dfu'$$

by commutativity. Hence

$$d(z - fu') = 0,$$

and $z - fu'$ is in the kernel of d. But $g(z - fu') = gz = z''$. This means that z'' is in the image of g_*, as desired. All the remaining cases of exactness will be left to the reader.

The original snake diagram may be completed by writing in the kernels and cokernels as follows (whence the name of the lemma):



§10. DIRECT AND INVERSE LIMITS

We return to limits, which we considered for groups in Chapter I. We now consider limits in other categories (rings, modules), and we point out that limits satisfy a universal property, in line with Chapter I, §11.

Let $I = \{i\}$ be a directed system of indices, defined in Chapter I, §10. Let $\mathfrak{A}$ be a category, and $\{A_i\}$ a family of objects in $\mathfrak{A}$. For each pair i, j such that

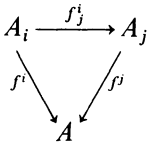
$i \leqq j$ assume given a morphism

$$f_j^i : A_i \to A_j$$

such that, whenever $i \leqq j \leqq k$, we have

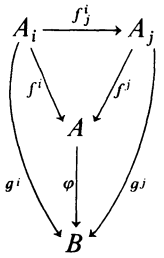
$$f_k^j \circ f_j^i = f_k^i \quad \text{and} \quad f_i^i = \mathrm{id}.$$

Such a family will be called a **directed family of morphisms**. A **direct limit** for the family $\{f_j^i\}$ is a universal object in the following category $\mathfrak{C}$. $\mathrm{Ob}(\mathfrak{C})$ consists of pairs $(A, (f^i))$ where $A \in \mathrm{Ob}(\mathfrak{A})$ and (f^i) is a family of morphisms $f^i : A_i \to A$, $i \in I$, such that for all $i \leqq j$ the following diagram is commutative:



(Universal of course means universally repelling.)

Thus if $(A, (f^i))$ is the direct limit, and if $(B, (g^i))$ is any object in the above category, then there exists a unique morphism $\varphi : A \to B$ which makes the following diagram commutative:



For simplicity, one usually writes

$$A = \varinjlim_i A_i,$$

omitting the f_j^i from the notation.

Theorem 10.1. *Direct limits exist in the category of abelian groups, or more generally in the category of modules over a ring.*

Proof. Let $\{M_i\}$ be a directed system of modules over a ring. Let M be their direct sum. Let N be the submodule generated by all elements

$$x_{ij} = (\ldots, 0, x, 0, \ldots, -f_j^i(x), 0, \ldots)$$

where, for a given pair of indices (i, j) with $j \geqq i$, x_{ij} has component x in M_i, $f_j^i(x)$ in M_j, and component 0 elsewhere. Then we leave to the reader the verification that the factor module M/N is a direct limit, where the maps of M_i into M/N are the natural ones arising from the composite homomorphism

$$M_i \to M \to M/N.$$

Example. Let X be a topological space, and let $x \in X$. The open neighborhoods of x form a directed system, by inclusion. Indeed, given two open neighborhoods U and V, then $U \cap V$ is also an open neighborhood contained in both U and V. In sheaf theory, one assigns to each U an abelian group $A(U)$ and to each pair $U \supset V$ a homomorphism $h_V^U\colon A(U) \to A(V)$ such that if $U \supset V \supset W$ then $h_W^V \circ h_V^U = h_W^U$. Then the family of such homomorphisms is a directed family. The direct limit

$$\varinjlim_U A(U)$$

is called the **stalk** at the point x. We shall give the formal definition of a sheaf of abelian groups in Chapter XX, §6. For further reading, I recommend at least two references. First, the self-contained short version of Chapter II in Hartshorne's *Algebraic Geometry*, Springer Verlag, 1977. (Do all the exercises of that section, concerning sheaves.) The section is only five pages long. Second, I recommend the treatment in Gunning's *Introduction to Holomorphic Functions of Several Variables*, Wadsworth and Brooks/Cole, 1990.

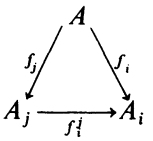
We now reverse the arrows to define inverse limits. We are again given a directed set I and a family of objects A_i. If $j \geqq i$ we are now given a morphism

$$f_i^j\colon A_j \to A_i$$

satisfying the relations

$$f_k^i \circ f_i^j = f_k^j \quad \text{and} \quad f_i^i = \mathrm{id},$$

if $j \geqq i$ and $i \geqq k$. As in the direct case, we can define a category of objects (A, f_i) with $f_i\colon A \to A_i$ such that for all i, j the following diagram is commutative:



A universal object in this category is called an **inverse limit** of the system (A_i, f_j^i).

As before, we often say that

$$A = \varprojlim_i A_i$$

is the inverse limit, omitting the f_j^i from the notation.

Theorem 10.2. *Inverse limits exist in the category of groups, in the category of modules over a ring, and also in the category of rings.*

Proof. Let $\{G_i\}$ be a directed family of groups, for instance, and let Γ be their inverse limit as defined in Chapter I, §10. Let $p_i: \Gamma \to G_i$ be the projection (defined as the restriction from the projection of the direct product, since Γ is a subgroup of $\prod G_i$). It is routine to verify that these data give an inverse limit in the category of groups. The same construction also applies to the category of rings and modules.

Example. Let p be a prime number. For $n \geqq m$ we have a canonical surjective ring homomorphism

$$f_m^n: \mathbf{Z}/p^n\mathbf{Z} \to \mathbf{Z}/p^m\mathbf{Z}.$$

The projective limit is called the ring of ***p*-adic integers**, and is denoted by $\mathbf{Z}_p$. For a consideration of this ring as a complete discrete valuation ring, see Exercise 17 and Chapter XII.

Let k be a field. The power series ring $k[[T]]$ in one variable may be viewed as the inverse limit of the factor polynomial rings $k[T]/(T^n)$, where for $n \geqq m$ we have the canonical ring homomorphism

$$f_m^n: k[T]/(T^n) \to k[T]/(T^m).$$

A similar remark applies to power series in several variables.

More generally, let R be a commutative ring and let J be a proper ideal. If $n \geqq m$ we have the canonical ring homomorphism

$$f_m^n: R/J^n \to R/J^m.$$

Let $\bar{R}_J = \lim R/J^n$ be the inverse limit. Then R has a natural homomorphism into $\bar{R}_J$. If R is a Noetherian local ring, then by Krull's theorem (Theorem 5.6 of Chapter X), one knows that $\cap J^n = \{0\}$, and so the natural homorphism of R in its completion is an embedding. This construction is applied especially when J is the maximal ideal. It gives an algebraic version of the notion of holomorphic functions for the following reason.

Let R be a commutative ring and J a proper ideal. Define a J-**Cauchy sequence** $\{x_n\}$ to be a sequence of elements of R satisfying the following condition. Given a positive integer k there exists N such that for all $n, m \geqq N$ we have $x_n - x_m \in J^k$. Define a **null sequence** to be a sequence for which given k there exists N such that for all $n \geqq N$ we have $x_n \in J^k$. Define addition and multipli-

cation of sequences termwise. Then the Cauchy sequences form a ring $\mathfrak{C}$, the null sequences form an ideal $\mathcal{N}$, and the factor ring $\mathfrak{C}/\mathcal{N}$ is called the J-**adic completion** of R. Prove these statements as an exercise, and also prove that there is a natural isomorphism

$$\mathfrak{C}/\mathcal{N} \approx \varprojlim R/J^n.$$

Thus the inverse limit $\varprojlim R/J^n$ is also called the J-adic completion. See Chapter XII for the completion in the context of absolute values on fields.

Examples. In certain situations one wants to determine whether there exist solutions of a system of a polynomial equation $f(X_1, \ldots, X_n) = 0$ with coefficients in a power series ring $k[[T]]$, say in one variable. One method is to consider the ring mod (T^N), in which case this equation amounts to a finite number of equations in the coefficients. A solution of $f(X) = 0$ is then viewed as an inverse limit of truncated solutions. For an early example of this method see [La 52], and for an extension to several variables [Ar 68].

[La 52] S. LANG, On quasi algebraic closure, *Ann of Math.* **55** (1952), pp. 373-390

[Ar 68] M. ARTIN, On the solutions of analytic equations, *Invent. Math.* **5** (1968), pp. 277-291

See also Chapter XII, §7.

In Iwasawa theory, one considers a sequence of Galois cyclic extensions K_n over a number field k of degree p^n with p prime, and with $K_n \subset K_{n+1}$. Let G_n be the Galois group of K_n over k. Then one takes the inverse limit of the group rings $(\mathbf{Z}/p^n\mathbf{Z})[G_n]$, following Iwasawa and Serre. Cf. my *Cyclotomic Fields*, Chapter 5. In such towers of fields, one can also consider the projective limits of the modules mentioned as examples at the end of §1. Specifically, consider the group of p^n-th roots of unity $\boldsymbol{\mu}_{p^n}$, and let $K_n = \mathbf{Q}(\boldsymbol{\mu}_{p^{n+1}})$, with $K_0 = \mathbf{Q}(\boldsymbol{\mu}_p)$. We let

$$T_p(\boldsymbol{\mu}) = \varprojlim \boldsymbol{\mu}_{p^n}$$

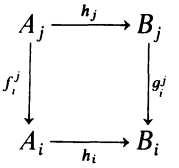
under the homomorphisms $\boldsymbol{\mu}_{p^{n+1}} \to \boldsymbol{\mu}_{p^n}$ given by $\zeta \mapsto \zeta^p$. Then $T_p(\boldsymbol{\mu})$ becomes a module for the projective limits of the group rings. Similarly, one can consider inverse limits for each one of the modules given in the examples at the end of §1. (See Exercise 18.) The determination of the structure of these inverse limits leads to fundamental problems in number theory and algebraic geometry.

After such examples from real life after basic algebra, we return to some general considerations about inverse limits.

Let $(A_i, f_i^j) = (A_i)$ and $(B_i, g_i^j) = (B_i)$ be two inverse systems of abelian groups indexed by the same indexing set. A **homomorphism** $(A_i) \to (B_i)$ is the obvious thing, namely a family of homomorphisms

$$h_i : A_i \to B_i$$

for each i which commute with the maps of the inverse systems:



A sequence

$$0 \to (A_i) \to (B_i) \to (C_i) \to 0$$

is said to be **exact** if the corresponding sequence of groups is exact for each i.

Let (A_n) be an inverse system of sets, indexed for simplicity by the positive integers, with connecting maps

$$u_{m,n} : A_m \to A_n \quad \text{for} \quad m \geqq n.$$

We say that this system satisfies the **Mittag-Leffler condition ML** if for each n, the decreasing sequence $u_{m,n}(A_m)$ $(m \geqq n)$ stabilizes, i.e. is constant for m sufficiently large. This condition is satisfied when $u_{m,n}$ is surjective for all m, n.

We note that trivially, the inverse limit functor is left exact, in the sense that given an exact sequence

$$0 \to (A_n) \to (B_n) \to (C_n) \to 0$$

then

$$0 \to \varprojlim A_n \to \varprojlim B_n \to \varprojlim C_n$$

is exact.

Proposition 10.3. *Assume that* (A_n) *satisfies* **ML**. *Given an exact sequence*

$$0 \to (A_n) \to (B_n) \xrightarrow{g} (C_n) \to 0$$

of inverse systems, then

$$0 \to \varprojlim A_n \to \varprojlim B_n \to \varprojlim C_n \to 0$$

is exact.

Proof. The only point is to prove the surjectivity on the right. Let (c_n) be an element of the inverse limit. Then each inverse image $g^{-1}(c_n)$ is a coset of A_n, so in bijection with A_n. These inverse images form an inverse system, and the **ML** condition on (A_n) implies **ML** on $(g^{-1}(c_n))$. Let S_n be the stable subset

$$S_n = \bigcap_{m \geqq n} u^B_{m,n}(g^{-1}(c_m)).$$

Then the connecting maps in the inverse system (S_n) are surjective, and so there is an element (b_n) in the inverse limit. It is immediate that g maps this element on the given (c_n), thereby concluding the proof of the Proposition.

Proposition 10.4. *Let (C_n) be an inverse system of abelian groups satisfying* **ML**, *and let $(u_{m,n})$ be the system of connecting maps. Then we have an exact sequence*

$$0 \to \varprojlim C_n \to \prod C_n \xrightarrow{1-u} \prod C_n \to 0.$$

Proof. For each positive integer N we have an exact sequence with a finite product

$$0 \to \lim_{1 \leqq n \leqq N} C_n \to \prod_{n=1}^{N} C_n \xrightarrow{1-u} \prod_{n=1}^{N} C_n \to 0.$$

The map u is the natural one, whose effect on a vector is

$$(0, \ldots, 0, c_m, 0, \ldots, 0) \mapsto (0, \ldots, 0, u_{m,m-1}c_m, 0, \ldots, 0).$$

One sees immediately that the sequence is exact. The infinite products are inverse limits taken over N. The hypothesis implies at once that **ML** is satisfied for the inverse limit on the left, and we can therefore apply Proposition 10.3 to conclude the proof.

EXERCISES

1. Let V be a vector space over a field K, and let U, W be subspaces. Show that
$$\dim U + \dim W = \dim(U + W) + \dim(U \cap W).$$
2. Generalize the dimension statement of Theorem 5.2 to free modules over a non zero commutative ring. [*Hint*: Recall how an analogous statement was proved for free abelian groups, and use a maximal ideal instead of a prime number.]
3. Let R be an entire ring containing a field k as a subring. Suppose that R is a finite dimensional vector space over k under the ring multiplication. Show that R is a field.
4. **Direct sums.**
   (a) Prove in detail that the conditions given in Proposition 3.2 for a sequence to split are equivalent. Show that a sequence $0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$ splits if and only if there exists a submodule N of M such that M is equal to the direct sum $\operatorname{Im} f \oplus N$, and that if this is the case, then N is isomorphic to M''. Complete all the details of the proof of Proposition 3.2.

(b) Let E and $E_i (i = 1, \ldots, m)$ be modules over a ring. Let $\varphi_i: E_i \to E$ and $\psi_i: E \to E_i$ be homomorphisms having the following properties:

$$\psi_i \circ \varphi_i = \mathrm{id}, \qquad \psi_i \circ \varphi_j = 0 \qquad \text{if } i \neq j,$$

$$\sum_{i=1}^{m} \varphi_i \circ \psi_i = \mathrm{id}.$$

Show that the map $x \mapsto (\psi_1 x, \ldots, \psi_m x)$ is an isomorphism of E onto the direct product of the E_i $(i = 1, \ldots, m)$, and that the map

$$(x_1, \ldots, x_m) \mapsto \varphi_1 x_1 + \cdots + \varphi_m x_m$$

is an isomorphism of this direct product onto E.

Conversely, if E is equal to a direct product (or direct sum) of submodules E_i $(i = 1, \ldots, m)$, if we let φ_i be the inclusion of E_i in E, and ψ_i the projection of E on E_i, then these maps satisfy the above-mentioned properties.

5. Let A be an additive subgroup of Euclidean space $\mathbf{R}^n$, and assume that in every bounded region of space, there is only a finite number of elements of A. Show that A is a free abelian group on $\leqq n$ generators. [*Hint*: Induction on the maximal number of linearly independent elements of A over $\mathbf{R}$. Let $v_1, \ldots, v_m$ be a maximal set of such elements, and let A_0 be the subgroup of A contained in the $\mathbf{R}$-space generated by $v_1, \ldots, v_{m-1}$. By induction, one may assume that any element of A_0 is a linear integral combination of $v_1, \ldots, v_{m-1}$. Let S be the subset of elements $v \in A$ of the form $v = a_1 v_1 + \cdots + a_m v_m$ with real coefficients a_i satisfying

$$0 \leqq a_i < 1 \qquad \text{if } i = 1, \ldots, m-1$$

$$0 \leqq a_m \leqq 1.$$

If v'_m is an element of S with the smallest $a_m \neq 0$, show that $\{v_1, \ldots, v_{m-1}, v'_m\}$ is a basis of A over $\mathbf{Z}$.]

Note. The above exercise is applied in algebraic number theory to show that the group of units in the ring of integers of a number field modulo torsion is isomorphic to a lattice in a Euclidean space. See Exercise 4 of Chapter VII.

6. (Artin-Tate). Let G be a finite group operating on a finite set S. For $w \in S$, denote $1 \cdot w$ by $[w]$, so that we have the direct sum

$$\mathbf{Z}\langle S \rangle = \sum_{w \in S} \mathbf{Z}[w].$$

Define an action of G on $\mathbf{Z}\langle S \rangle$ by defining $\sigma[w] = [\sigma w]$ (for $w \in S$), and extending σ to $\mathbf{Z}\langle S \rangle$ by linearity. Let M be a subgroup of $\mathbf{Z}\langle S \rangle$ of rank $\#[S]$. Show that M has a $\mathbf{Z}$-basis $\{y_w\}_{w \in S}$ such that $\sigma y_w = y_{\sigma w}$ for all $w \in S$. (Cf. my *Algebraic Number Theory*, Chapter IX, §4, Theorem 1.)

7. Let M be a finitely generated abelian group. By a **seminorm** on M we mean a real-valued function $v \mapsto |v|$ satisfying the following properties:

$$|v| \geqq 0 \text{ for all } v \in M;$$
$$|nv| = |n|\,|v| \text{ for } n \in \mathbf{Z};$$
$$|v + w| \leqq |v| + |w| \text{ for all } v, w \in M.$$

By the **kernel** of the seminorm we mean the subset of elements v such that $|v| = 0$.

(a) Let M_0 be the kernel. Show that M_0 is a subgroup. If $M_0 = \{0\}$, then the seminorm is called a **norm**.

(b) Assume that M has rank r. Let $v_1, \ldots, v_r \in M$ be linearly independent over $\mathbf{Z}$ mod M_0. Prove that there exists a basis $\{w_1, \ldots, w_r\}$ of M/M_0 such that

$$|w_i| \leqq \sum_{j=1}^{i} |v_j|.$$

[*Hint*: An explicit version of the proof of Theorem 7.8 gives the result. Without loss of generality, we can asume $M_0 = \{0\}$. Let $M_1 = \langle v_1, \ldots, v_r \rangle$. Let d be the exponent of M/M_1. Then dM has a finite index in M_1. Let $n_{j,j}$ be the smallest positive integer such that there exist integers $n_{j,1}, \ldots, n_{j,j-1}$ satisfying

$$n_{j,1}v_1 + \cdots + n_{j,j}v_j = dw_j \text{ for some } w_j \in M.$$

Without loss of generality we may assume $0 \leqq n_{j,k} \leqq d - 1$. Then the elements $w_1, \ldots, w_r$ form the desired basis.]

8. Consider the multiplicative group $\mathbf{Q}^*$ of non-zero rational numbers. For a non-zero rational number $x = a/b$ with $a, b \in \mathbf{Z}$ and $(a, b) = 1$, define the **height**

$$h(x) = \log \max(|a|, |b|).$$

(a) Show that h defines a seminorm on $\mathbf{Q}^*$, whose kernel consists of ± 1 (the torsion group).

(b) Let M_1 be a finitely generated subgroup of $\mathbf{Q}^*$, generated by rational numbers $x_1, \ldots, x_m$. Let M be the subgroup of $\mathbf{Q}^*$ consisting of those elements x such that $x^s \in M_1$ for some positive integer s. Show that M is finitely generated, and using Exercise 7, find a bound for the seminorm of a set of generators of M in terms of the seminorms of $x_1, \ldots, x_m$.

Note. The above two exercises are applied in questions of diophantine approximation. See my Diophantine approximation on toruses, *Am. J. Math.* **86** (1964), pp. 521-533, and the discussion and references I give in *Encyclopedia of Mathematical Sciences, Number Theory III*, Springer Verlag, 1991, pp. 240-243.

Localization

9. (a) Let A be a commutative ring and let M be an A-module. Let S be a multiplicative subset of A. Define $S^{-1}M$ in a manner analogous to the one we used to define $S^{-1}A$, and show that $S^{-1}M$ is an $S^{-1}A$-module.

(b) If $0 \to M' \to M \to M'' \to 0$ is an exact sequence, show that the sequence $0 \to S^{-1}M' \to S^{-1}M \to S^{-1}M'' \to 0$ is exact.

10. (a) If $\mathfrak{p}$ is a prime ideal, and $S = A - \mathfrak{p}$ is the complement of $\mathfrak{p}$ in the ring A, then $S^{-1}M$ is denoted by $M_{\mathfrak{p}}$. Show that the natural map

$$M \to \prod M_{\mathfrak{p}}$$

of a module M into the direct product of all localizations $M_{\mathfrak{p}}$ where $\mathfrak{p}$ ranges over all *maximal* ideals, is injective.

(b) Show that a sequence $0 \to M' \to M \to M'' \to 0$ is exact if and only if the sequence $0 \to M'_{\mathfrak{p}} \to M_{\mathfrak{p}} \to M''_{\mathfrak{p}} \to 0$ is exact for all primes $\mathfrak{p}$.

(c) Let A be an entire ring and let M be a torsion-free module. For each prime $\mathfrak{p}$ of A show that the natural map $M \to M_{\mathfrak{p}}$ is injective. In particular $A \to A_{\mathfrak{p}}$ is injective, but you can see that directly from the imbedding of A in its quotient field K.

Projective modules over Dedekind rings

For the next exercise we assume you have done the exercises on Dedekind rings in the preceding chapter. We shall see that for such rings, some parts of their module theory can be reduced to the case of principal rings by localization. We let $\mathfrak{o}$ be a Dedekind ring and K its quotient field.

11. Let M be a finitely generated torsion-free module over $\mathfrak{o}$. Prove that M is projective. [*Hint*: Given a prime ideal $\mathfrak{p}$, the localized module $M_{\mathfrak{p}}$ is finitely generated torsion-free over $\mathfrak{o}_{\mathfrak{p}}$, which is principal. Then $M_{\mathfrak{p}}$ is projective, so if F is finite free over $\mathfrak{o}$, and $f\colon F \to M$ is a surjective homomorphism, then $f_{\mathfrak{p}}\colon F_{\mathfrak{p}} \to M_{\mathfrak{p}}$ has a splitting $g_{\mathfrak{p}}\colon M_{\mathfrak{p}} \to F_{\mathfrak{p}}$, such that $f_{\mathfrak{p}} \circ g_{\mathfrak{p}} = \mathrm{id}_{M_{\mathfrak{p}}}$. There exists $c_{\mathfrak{p}} \in \mathfrak{o}$ such that $c_{\mathfrak{p}} \notin \mathfrak{p}$ and $c_{\mathfrak{p}} g_{\mathfrak{p}}(M) \subset F$. The family $\{c_{\mathfrak{p}}\}$ generates the unit ideal $\mathfrak{o}$ (why?), so there is a finite number of elements $c_{\mathfrak{p}_i}$ and elements $x_i \in \mathfrak{o}$ such that $\sum x_i c_{\mathfrak{p}_i} = 1$. Let

$$g = \sum x_i c_{\mathfrak{p}_i} g_{\mathfrak{p}_i}.$$

Then show that $g\colon M \to F$ gives a homomorphism such that $f \circ g = \mathrm{id}_M$.]

12. (a) Let $\mathfrak{a}, \mathfrak{b}$ be ideals. Show that there is an isomorphism of $\mathfrak{o}$-modules

$$\mathfrak{a} \oplus \mathfrak{b} \xrightarrow{\approx} \mathfrak{o} \oplus \mathfrak{a}\mathfrak{b}$$

[*Hint*: First do this when $\mathfrak{a}, \mathfrak{b}$ are relatively prime. Consider the homomorphism $\mathfrak{a} \oplus \mathfrak{b} \to \mathfrak{a} + \mathfrak{b}$, and use Exercise 10. Reduce the general case to the relatively prime case by using Exercise 19 of Chapter II.]

(b) Let $\mathfrak{a}, \mathfrak{b}$ be fractional ideals, and let $f\colon \mathfrak{a} \to \mathfrak{b}$ be an isomorphism (of $\mathfrak{o}$-modules, of course). Then f has an extension to a K-linear map $f_K\colon K \to K$. Let $c = f_K(1)$. Show that $\mathfrak{b} = c\mathfrak{a}$ and that f is given by the mapping $m_c\colon x \to cx$ (multiplication by c).

(c) Let $\mathfrak{a}$ be a fractional ideal. For each $b \in \mathfrak{a}^{-1}$ the map $m_b\colon \mathfrak{a} \to \mathfrak{o}$ is an element of the dual $\mathfrak{a}^\vee$. Show that $\mathfrak{a}^{-1} = \mathfrak{a}^\vee = \mathrm{Hom}_{\mathfrak{o}}(\mathfrak{a}, \mathfrak{o})$ under this map, and so $\mathfrak{a}^{\vee\vee} = \mathfrak{a}$.

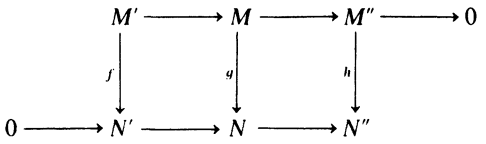
13. (a) Let M be a projective finite module over the Dedekind ring $\mathfrak{o}$. Show that there exist free modules F and F' such that $F \supset M \supset F'$, and F, F' have the same rank, which is called the **rank** of M.

(b) Prove that there exists a basis $\{e_1, \ldots, e_n\}$ of F and ideals $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ such that $M = \mathfrak{a}_1 e_1 + \cdots + \mathfrak{a}_n e_n$, or in other words, $M \approx \oplus\, \mathfrak{a}_i$.

(c) Prove that $M \approx \mathfrak{o}^{n-1} \oplus \mathfrak{a}$ for some ideal $\mathfrak{a}$, and that the association $M \mapsto \mathfrak{a}$ induces an isomorphism of $K_0(\mathfrak{o})$ with the group of ideal classes $\mathrm{Pic}(\mathfrak{o})$. (The group $K_0(\mathfrak{o})$ is the group of equivalence classes of projective modules defined at the end of §4.)

A few snakes

14. Consider a commutative diagram of R-modules and homomorphisms such that each row is exact:



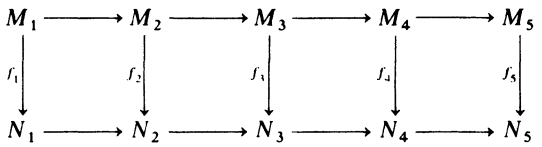
Prove:

(a) If f, h are monomorphisms then g is a monomorphism.

(b) If f, h are surjective, then g is surjective.

(c) Assume in addition that $0 \to M' \to M$ is exact and that $N \to N'' \to 0$ is exact. Prove that if any two of f, g, h are isomorphisms, then so is the third. [*Hint*: Use the snake lemma.]

15. **The five lemma.** Consider a commutative diagram of R-modules and homomorphisms such that each row is exact:



Prove:

(a) If f_1 is surjective and f_2, f_4 are monomorphisms, then f_3 is a monomorphism.

(b) If f_5 is a monomorphism and f_2, f_4 are surjective, then f_3 is surjective. [*Hint*: Use the snake lemma.]

Inverse limits

16. Prove that the inverse limit of a system of simple groups in which the homomorphisms are surjective is either the trivial group, or a simple group.

17. (a) Let n range over the positive integers and let p be a prime number. Show that the abelian groups $A_n = \mathbf{Z}/p^n\mathbf{Z}$ form an inverse system under the canonical homomorphism if $n \geqq m$. Let $\mathbf{Z}_p$ be its inverse limit. Show that $\mathbf{Z}_p$ maps surjectively on each $\mathbf{Z}/p^n\mathbf{Z}$; that $\mathbf{Z}_p$ has no divisors of 0, and has a unique maximal ideal generated by p. Show that $\mathbf{Z}_p$ is factorial, with only one prime, namely p itself.

(b) Next consider all non zero ideals of $\mathbf{Z}$ as forming a directed system, by divisibility. Prove that

$$\varprojlim_{(a)} \mathbf{Z}/(a) = \prod_p \mathbf{Z}_p,$$

where the limit is taken over all non zero ideals (a), and the product is taken over all primes p.

18. (a) Let $\{A_n\}$ be an inversely directed sequence of commutative rings, and let $\{M_n\}$ be an inversely directed sequence of modules, M_n being a module over A_n such that the following diagram is commutative:

$$\begin{array}{ccccc} A_{n+1} & \times & M_{n+1} & \to & M_{n+1} \\ \downarrow & & \downarrow & & \downarrow \\ A_n & \times & M_n & \to & M_n \end{array}$$

The vertical maps are the homomorphisms of the directed sequence, and the horizontal maps give the operation of the ring on the module. Show that $\varprojlim M_n$ is a module over $\varprojlim A_n$.

(b) Let M be a p-divisible group. Show that $T_p(A)$ is a module over $\mathbf{Z}_p$.

(c) Let M, N be p-divisible groups. Show that $T_p(M \oplus N) = T_p(M) \oplus T_p(N)$, as modules over $\mathbf{Z}_p$.

Direct limits

19. Let (A_i, f_j^i) be a directed family of modules. Let $a_k \in A_k$ for some k, and suppose that the image of a_k in the direct limit A is 0. Show that there exists some index $j \geqq k$ such that $f_j^k(a_k) = 0$. In other words, whether some element in some group A_i vanishes in the direct limit can already be seen within the original data. One way to see this is to use the construction of Theorem 10.1.

20. Let I, J be two directed sets, and give the product $I \times J$ the obvious ordering that $(i, j) \leqq (i', j')$ if $i \leqq i'$ and $j \leqq j'$. Let A_{ij} be a family of abelian groups, with homomorphisms indexed by $I \times J$, and forming a directed family. Show that the direct limits

$$\varinjlim_i \varinjlim_j A_{ij} \quad \text{and} \quad \varinjlim_j \varinjlim_i A_{ij}$$

exist and are isomorphic in a natural way. State and prove the same result for inverse limits.

21. Let (M_i', f_j^i), (M_i, g_j^i) be directed systems of modules over a ring. By a **homomorphism**

$$(M_i') \xrightarrow{u} (M_i)$$

one means a family of homomorphisms $u_i : M_i' \to M_i$ for each i which commute with the f_j^i, g_j^i. Suppose we are given an exact sequence

$$0 \to (M_i') \xrightarrow{u} (M_i) \xrightarrow{v} (M_i'') \to 0$$

of directed systems, meaning that for each i, the sequence

$$0 \to M_i' \to M_i \to M_i'' \to 0$$

is exact. Show that the direct limit preserves exactness, that is

$$0 \to \varinjlim M_i' \to \varinjlim M_i \to \varinjlim M_i'' \to 0$$

is exact.

22. (a) Let $\{M_i\}$ be a family of modules over a ring. For any module N show that

$$\mathrm{Hom}(\bigoplus M_i, N) = \prod \mathrm{Hom}(M_i, N)$$

(b) Show that

$$\mathrm{Hom}(N, \prod M_i) = \prod \mathrm{Hom}(N, M_i).$$

23. Let $\{M_i\}$ be a directed family of modules over a ring. For any module N show that

$$\varinjlim \mathrm{Hom}(N, M_i) = \mathrm{Hom}(N, \varinjlim M_i)$$

24. Show that any module is a direct limit of finitely generated submodules.

A module M is called **finitely presented** if there is an exact sequence

$$F_1 \to F_0 \to M \to 0$$

where F_0, F_1 are free with finite bases. The image of F_1 in F_0 is said to be the submodule of **relations**, among the free basis elements of F_0.

25. Show that any module is a direct limit of finitely presented modules (not necessarily submodules). In other words, given M, there exists a directed system $\{M_i, f_j^i\}$ with M_i finitely presented for all i such that

$$M \approx \varinjlim M_i.$$

[*Hint*: Any finitely generated submodule is such a direct limit, since an infinitely generated module of relations can be viewed as a limit of finitely generated modules of relations. Make this precise to get a proof.]

26. Let E be a module over a ring. Let $\{M_i\}$ be a directed family of modules. If E is finitely generated, show that the natural homomorphism

$$\varinjlim \mathrm{Hom}(E, M_i) \to \mathrm{Hom}(E, \varinjlim M_i)$$

is injective. If E is finitely presented, show that this homomorphism is an isomorphism. *Hint*: First prove the statements when E is free with finite basis. Then, say E is finitely presented by an exact sequence $F_1 \to F_0 \to E \to 0$. Consider the diagram:

$$\begin{array}{ccccccc}
0 & \longrightarrow & \varinjlim \mathrm{Hom}(E, M_i) & \longrightarrow & \varinjlim \mathrm{Hom}(F_0, M_i) & \longrightarrow & \varinjlim \mathrm{Hom}(F_1, M_i) \\
 & & \downarrow & & \downarrow & & \downarrow \\
0 & \longrightarrow & \mathrm{Hom}(E, \varinjlim M_i) & \longrightarrow & \mathrm{Hom}(F_0, \varinjlim M_i) & \longrightarrow & \mathrm{Hom}(F_1, \varinjlim M_i)
\end{array}$$

Graded Algebras

Let A be an algebra over a field k. By a **filtration** of A we mean a sequence of k-vector spaces A_i $(i = 0, 1, \ldots)$ such that

$$A_0 \subset A_1 \subset A_2 \subset \cdots \qquad \text{and} \qquad \bigcup A_i = A,$$

and $A_iA_j \subset A_{i+j}$ for all $i, j \geqq 0$. We then call A a filtered algebra. Let R be an algebra. We say that R is **graded** if R is a direct sum $R = \bigoplus R_i$ of subspaces such that $R_iR_j \subset R_{i+j}$ for all $i, j \geqq 0$.

27. Let A be a filtered algebra. Define R_i for $i \geqq 0$ by $R_i = A_i/A_{i-1}$. By definition, $A_{-1} = \{0\}$. Let $R = \bigoplus R_i$, and $R_i = \mathrm{gr}_i(A)$. Define a natural product on R making R into a graded algebra, denoted by $\mathrm{gr}(A)$, and called the **associated graded algebra**.

28. Let A, B be filtered algebras, $A = \bigcup A_i$ and $B = \bigcup B_i$. Let $L: A \to B$ be a k-linear map preserving the filtration, that is $L(A_i) \subset B_i$ for all i, and $L(ca) = L(c)L(a)$ for $c \in k$ and $a \in A_i$ for all i.
    (a) Show that L induces a k-linear map
    $$\mathrm{gr}_i(L): \mathrm{gr}_i(A) \to \mathrm{gr}_i(B) \qquad \text{for all } i.$$
    (b) Suppose that $\mathrm{gr}_i(L)$ is an isomorphism for all i. Show that L is a k-linear isomorphism.

29. Suppose k has characteristic 0. Let $\mathfrak{n}$ be the set of all strictly upper triangular matrices of a given size $n \times n$ over k.
    (a) For a given matrix $X \in \mathfrak{n}$, let $D_1(X), \ldots, D_n(X)$ be its diagonals, so $D_1 = D_1(X)$ is the main diagonal, and is 0 by the definition of $\mathfrak{n}$. Let $\mathfrak{n}_i$ be the subset of $\mathfrak{n}$ consisting of those matrices whose diagonals $D_1, \ldots, D_{n-i}$ are 0. Thus $\mathfrak{n}_0 = \{0\}$, $\mathfrak{n}_1$ consists of all matrices whose components are 0 except possibly for x_{nn}; $\mathfrak{n}_2$ consists of all matrices whose components are 0 except possibly those in the last two diagonals; and so forth. Show that each $\mathfrak{n}_i$ is an algebra, and its elements are nilpotent (in fact the $(i+1)$-th power of its elements is 0).
    (b) Let U be the set of elements $I + X$ with $X \in \mathfrak{n}$. Show that U is a multiplicative group.
    (c) Let exp be the exponential series defined as usual. Show that exp defines a polynomial function on $\mathfrak{n}$ (all but a finite number of terms are 0 when evaluated on a nilpotent matrix), and establishes a bijection
    $$\exp: \mathfrak{n} \to U.$$
    Show that the inverse is given by the standard log series.

CHAPTER IV

Polynomials

This chapter provides a continuation of Chapter II, §3. We prove standard properties of polynomials. Most readers will be acquainted with some of these properties, especially at the beginning for polynomials in one variable. However, one of our purposes is to show that some of these properties also hold over a commutative ring when properly formulated. The Gauss lemma and the reduction criterion for irreducibility will show the importance of working over rings. Chapter IX will give examples of the importance of working over the integers **Z** themselves to get universal relations. It happens that certain statements of algebra are universally true. To prove them, one proves them first for elements of a polynomial ring over **Z**, and then one obtains the statement in arbitrary fields (or commutative rings as the case may be) by specialization. The Cayley–Hamilton theorem of Chapter XV, for instance, can be proved in that way.

The last section on power series shows that the basic properties of polynomial rings can be formulated so as to hold for power series rings. I conclude this section with several examples showing the importance of power series in various parts of mathematics.

§1. BASIC PROPERTIES FOR POLYNOMIALS IN ONE VARIABLE

We start with the Euclidean algorithm.

Theorem 1.1. *Let A be a commutative ring, let $f, g \in A[X]$ be polynomials in one variable, of degrees $\geqq 0$, and assume that the leading*

S. Lang, *Algebra*
© Springer Science+Business Media LLC 2002

coefficient of g is a unit in A. Then there exist unique polynomials $q, r \in A[X]$ such that

$$f = gq + r$$

and $\deg r < \deg g$.

Proof. Write

$$f(X) = a_n X^n + \cdots + a_0,$$
$$g(X) = b_d X^d + \cdots + b_0,$$

where $n = \deg f$, $d = \deg g$ so that a_n, $b_d \neq 0$ and b_d is a unit in A. We use induction on n.

If $n = 0$, and $\deg g > \deg f$, we let $q = 0$, $r = f$. If $\deg g = \deg f = 0$, then we let $r = 0$ and $q = a_n b_d^{-1}$.

Assume the theorem proved for polynomials of degree $< n$ (with $n > 0$). We may assume $\deg g \leqq \deg f$ (otherwise, take $q = 0$ and $r = f$). Then

$$f(X) = a_n b_d^{-1} X^{n-d} g(X) + f_1(X),$$

where $f_1(X)$ has degree $< n$. By induction, we can find q_1, r such that

$$f(X) = a_n b_d^{-1} X^{n-d} g(X) + q_1(X) g(X) + r(X)$$

and $\deg r < \deg g$. Then we let

$$q(X) = a_n b_d^{-1} X^{n-d} + q_1(X)$$

to conclude the proof of existence for q, r.

As for uniqueness, suppose that

$$f = q_1 g + r_1 = q_2 g + r_2$$

with $\deg r_1 < \deg g$ and $\deg r_2 < \deg g$. Subtracting yields

$$(q_1 - q_2) g = r_2 - r_1.$$

Since the leading coefficient of g is assumed to be a unit, we have

$$\deg(q_1 - q_2) g = \deg(q_1 - q_2) + \deg g.$$

Since $\deg(r_2 - r_1) < \deg g$, this relation can hold only if $q_1 - q_2 = 0$, i.e. $q_1 = q_2$, and hence finally $r_1 = r_2$ as was to be shown.

Theorem 1.2. *Let k be a field. Then the polynomial ring in one variable $k[X]$ is principal.*

Proof. Let $\mathfrak{a}$ be an ideal of $k[X]$, and assume $\mathfrak{a} \neq 0$. Let g be an element of $\mathfrak{a}$ of smallest degree $\geqq 0$. Let f be any element of $\mathfrak{a}$ such that $f \neq 0$. By the Euclidean algorithm we can find $q, r \in k[X]$ such that

$$f = qg + r$$

and $\deg r < \deg g$. But $r = f - qg$, whence r is in $\mathfrak{a}$. Since g had minimal degree $\geqq 0$ it follows that $r = 0$, hence that $\mathfrak{a}$ consists of all polynomials qg (with $q \in k[X]$). This proves our theorem. By Theorem 5.2 of Chapter II we get:

Corollary 1.3. *The ring $k[X]$ is factorial.*

If k is a field then every non-zero element of k is a unit in k, and one sees immediately that the units of $k[X]$ are simply the units of k. (No polynomial of degree $\geqq 1$ can be a unit because of the addition formula for the degree of a product.)

A polynomial $f(X) \in k[X]$ is called **irreducible** if it has degree $\geqq 1$, and if one cannot write $f(X)$ as a product

$$f(X) = g(X)h(X)$$

with $g, h \in k[X]$, and both $g, h \notin k$. Elements of k are usually called **constant polynomials**, so we can also say that in such a factorization, one of g or h must be constant. A polynomial is called **monic** if it has leading coefficient 1.

Let A be a commutative ring and $f(X)$ a polynomial in $A[X]$. Let A be a subring of B. An element $b \in B$ is called a **root** or a **zero** of f in B if $f(b) = 0$. Similarly, if (X) is an n-tuple of variables, an n-tuple (b) is called a zero of f if $f(b) = 0$.

Theorem 1.4. *Let k be a field and f a polynomial in one variable X in $k[X]$, of degree $n \geqq 0$. Then f has at most n roots in k, and if a is a root of f in k, then $X - a$ divides $f(X)$.*

Proof. Suppose $f(a) = 0$. Find q, r such that

$$f(X) = q(X)(X - a) + r(X)$$

and $\deg r < 1$. Then

$$0 = f(a) = r(a).$$

Since $r = 0$ or r is a non-zero constant, we must have $r = 0$, whence $X - a$ divides $f(X)$. If $a_1, \ldots, a_m$ are distinct roots of f in k, then inductively we see that the product

$$(X - a_1) \cdots (X - a_m)$$

divides $f(X)$, whence $m \leqq n$, thereby proving the theorem. The next corollaries give applications of Theorem 1.4 to polynomial functions.

Corollary 1.5. *Let k be a field and T an infinite subset of k. Let $f(X) \in k[X]$ be a polynomial in one variable. If $f(a) = 0$ for all $a \in T$, then $f = 0$, i.e. f induces the zero function.*

Corollary 1.6. *Let k be a field, and let $S_1, \ldots, S_n$ be infinite subsets of k. Let $f(X_1, \ldots, X_n)$ be a polynomial in n variables over k. If $f(a_1, \ldots, a_n) = 0$ for all $a_i \in S_i$ $(i = 1, \ldots, n)$, then $f = 0$.*

Proof. By induction. We have just seen the result is true for one variable. Let $n \geqq 2$, and write

$$f(X_1, \ldots, X_n) = \sum_j f_j(X_1, \ldots, X_{n-1}) X_n^j$$

as a polynomial in X_n with coefficients in $k[X_1, \ldots, X_{n-1}]$. If there exists

$$(b_1, \ldots, b_{n-1}) \in S_1 \times \cdots \times S_{n-1}$$

such that for some j we have $f_j(b_1, \ldots, b_{n-1}) \neq 0$, then

$$f(b_1, \ldots, b_{n-1}, X_n)$$

is a non-zero polynomial in $k[X_n]$ which takes on the value 0 for the infinite set of elements S_n. This is impossible. Hence f_j induces the zero function on $S_1 \times \cdots \times S_{n-1}$ for all j, and by induction we have $f_j = 0$ for all j. Hence $f = 0$, as was to be shown.

Corollary 1.7. *Let k be an infinite field and f a polynomial in n variables over k. If f induces the zero function on $k^{(n)}$, then $f = 0$.*

We shall now consider the case of finite fields. Let k be a finite field with q elements. Let $f(X_1, \ldots, X_n)$ be a polynomial in n variables over k. Write

$$f(X_1, \ldots, X_n) = \sum a_{(\nu)} X_1^{\nu_1} \cdots X_n^{\nu_n}.$$

If $a_{(\nu)} \neq 0$, we recall that the monomial $M_{(\nu)}(X)$ **occurs** in f. Suppose this is the case, and that in this monomial $M_{(\nu)}(X)$, some variable X_i occurs with an exponent $\nu_i \geqq q$. We can write

$$X_i^{\nu_i} = X_i^{q+\mu}, \qquad \mu = \text{integer} \geqq 0.$$

If we now replace $X_i^{\nu_i}$ by $X_i^{\mu+1}$ in this monomial, then we obtain a new polynomial which gives rise to the same function as f. The degree of this new polynomial is at most equal to the degree of f.

Performing the above operation a finite number of times, for all the monomials occurring in f and all the variables $X_1, \ldots, X_n$ we obtain some polynomial f^* giving rise to the same function as f, but whose degree in each variable is $< q$.

Corollary 1.8. *Let k be a finite field with q elements. Let f be a polynomial in n variables over k such that the degree of f in each variable is $< q$. If f induces the zero function on $k^{(n)}$, then $f = 0$.*

Proof. By induction. If $n = 1$, then the degree of f is $< q$, and hence f cannot have q roots unless it is 0. The inductive step is carried out just as we did for the proof of Corollary 1.6 above.

Let f be a polynomial in n variables over the finite field k. A polynomial g whose degree in each variable is $< q$ will be said to be **reduced**. We have shown above that there exists a reduced polynomial f^* which gives the same function as f on $k^{(n)}$. Theorem 1.8 now shows that *this reduced polynomial is unique.* Indeed, if g_1, g_2 are reduced polynomials giving the same function, then $g_1 - g_2$ is reduced and gives the zero function. Hence $g_1 - g_2 = 0$ and $g_1 = g_2$.

We shall give one more application of Theorem 1.4. Let k be a field. By a **multiplicative subgroup** of k we shall mean a subgroup of the group k^* (non-zero elements of k).

Theorem 1.9. *Let k be a field and let U be a finite multiplicative subgroup of k. Then U is cyclic.*

Proof. Write U as a product of subgroups $U(p)$ for each prime p, where $U(p)$ is a p-group. By Proposition 4.3(v) of Chapter I, it will suffice to prove that $U(p)$ is cyclic for each p. Let a be an element of $U(p)$ of maximal period p^r for some integer r. Then $x^{p^r} = 1$ for every element $x \in U(p)$, and hence all elements of $U(p)$ are roots of the polynomial

$$X^{p^r} - 1.$$

The cyclic group generated by a has p^r elements. If this cyclic group is not equal to $U(p)$, then our polynomial has more than p^r roots, which is impossible. Hence a generates $U(p)$, and our theorem is proved.

Corollary 1.10. *If k is a finite field, then k^* is cyclic.*

An element ζ in a field k such that there exists an integer $n \geqq 1$ such that $\zeta^n = 1$ is called a **root of unity**, or more precisely an n-th root of unity. Thus the set of n-th roots of unity is the set of roots of the polynomial $X^n - 1$. There are at most n such roots, and they obviously form a group, which is

cyclic by Theorem 1.9. We shall study roots of unity in greater detail later. A generator for the group of n-th roots of unity is called a **primitive** n-th root of unity. For example, in the complex numbers, $e^{2\pi i/n}$ is a primitive n-th root of unity, and the n-th roots of unity are of type $e^{2\pi iv/n}$ with $1 \leqq v \leqq n$.

The group of roots of unity is denoted by $\boldsymbol{\mu}$. The group of roots of unity in a field K is denoted by $\boldsymbol{\mu}(K)$.

A field k is said to be **algebraically closed** if every polynomial in $k[X]$ of degree $\geqq 1$ has a root in k. In books on analysis, it is proved that the complex numbers are algebraically closed. In Chapter V we shall prove that a field k is always contained in some algebraically closed field. If k is algebraically closed then the irreducible polynomials in $k[X]$ are the polynomials of degree 1. In such a case, the unique factorization of a polynomial f of degree $\geqq 0$ can be written in the form

$$f(X) = c \prod_{i=1}^{r} (X - \alpha_i)^{m_i}$$

with $c \in k$, $c \neq 0$ and distinct roots $\alpha_1, \ldots, \alpha_r$. We next develop a test when $m_i > 1$.

Let A be a commutative ring. We define a map

$$D\colon A[X] \to A[X]$$

of the polynomial ring into itself. If $f(X) = a_n X^n + \cdots + a_0$ with $a_i \in A$, we define the **derivative**

$$Df(X) = f'(X) = \sum_{v=1}^{n} v a_v X^{v-1} = n a_n X^{n-1} + \cdots + a_1.$$

One verifies easily that if f, g are polynomials in $A[X]$, then

$$(f + g)' = f' + g', \qquad (fg)' = f'g + fg',$$

and if $a \in A$, then

$$(af)' = af'.$$

Let K be a field and f a non-zero polynomial in $K[X]$. Let a be a root of f in K. We can write

$$f(X) = (X - a)^m g(X)$$

with some polynomial $g(X)$ relatively prime to $X - a$ (and hence such that $g(a) \neq 0$). We call m the **multiplicity** of a in f, and say that a is a **multiple root** if $m > 1$.

Proposition 1.11. *Let K, f be as above. The element a of K is a multiple root of f if and only if it is a root and $f'(a) = 0$.*

Proof. Factoring f as above, we get

$$f'(X) = (X - a)^m g'(X) + m(X - a)^{m-1} g(X).$$

If $m > 1$, then obviously $f'(a) = 0$. Conversely, if $m = 1$ then

$$f'(X) = (X - a)g'(X) + g(X),$$

whence $f'(a) = g(a) \neq 0$. Hence if $f'(a) = 0$ we must have $m > 1$, as desired.

Proposition 1.12. *Let $f \in K[X]$. If K has characteristic 0, and f has degree $\geqq 1$, then $f' \neq 0$. Let K have characteristic $p > 0$ and f have degree $\geqq 1$. Then $f' = 0$ if and only if, in the expression for $f(X)$ given by*

$$f(X) = \sum_{v=0}^{n} a_v X^v,$$

p divides each integer v such that $a_v \neq 0$.

Proof. If K has characteristic 0, then the derivative of a monomial $a_v X^v$ such that $v \geqq 1$ and $a_v \neq 0$ is not zero since it is $v a_v X^{v-1}$. If K has characteristic $p > 0$, then the derivative of such a monomial is 0 if and only if $p | v$, as contended.

Let K have characteristic $p > 0$, and let f be written as above, and be such that $f'(X) = 0$. Then one can write

$$f(X) = \sum_{\mu=0}^{d} b_\mu X^{p\mu}$$

with $b_\mu \in K$.

Since the binomial coefficients $\binom{p}{v}$ are divisible by p for $1 \leqq v \leqq p - 1$ we see that if K has characteristic p, then for $a, b \in K$ we have

$$(a + b)^p = a^p + b^p.$$

Since obviously $(ab)^p = a^p b^p$, the map

$$x \mapsto x^p$$

is a homomorphism of K into itself, which has trivial kernel, hence is injective. Iterating, we conclude that for each integer $r \geqq 1$, the map $x \mapsto x^{p^r}$

is an endomorphism of K, called the **Frobenius endomorphism**. Inductively, if $c_1, \ldots, c_n$ are elements of K, then

$$(c_1 + \cdots + c_n)^p = c_1^p + \cdots + c_n^p.$$

Applying these remarks to polynomials, we see that for any element $a \in K$ we have

$$(X - a)^{p^r} = X^{p^r} - a^{p^r}.$$

If $c \in K$ and the polynomial

$$X^{p^r} - c$$

has one root a in K, then $a^{p^r} = c$ and

$$X^{p^r} - c = (X - a)^{p^r}.$$

Hence our polynomial has precisely one root, of multiplicity p^r. For instance, $(X - 1)^{p^r} = X^{p^r} - 1$.

§2. POLYNOMIALS OVER A FACTORIAL RING

Let A be a factorial ring, and K its quotient field. Let $a \in K$, $a \neq 0$. We can write a as a quotient of elements in A, having no prime factor in common. If p is a prime element of A, then we can write

$$a = p^r b,$$

where $b \in K$, r is an integer, and p does not divide the numerator or denominator of b. Using the unique factorization in A, we see at once that r is uniquely determined by a, and we call r the **order of a at p** (and write $r = \operatorname{ord}_p a$). If $a = 0$, we define its order at p to be ∞.

If $a, a' \in K$ and $aa' \neq 0$, then

$$\operatorname{ord}_p(aa') = \operatorname{ord}_p a + \operatorname{ord}_p a'.$$

This is obvious.

Let $f(X) \in K[X]$ be a polynomial in one variable, written

$$f(X) = a_0 + a_1 X + \cdots + a_n X^n.$$

If $f = 0$, we define $\operatorname{ord}_p f$ to be ∞. If $f \neq 0$, we define $\operatorname{ord}_p f$ to be

$$\operatorname{ord}_p f = \min \operatorname{ord}_p a_i,$$

the minimum being taken over all those i such that $a_i \neq 0$.

If $r = \operatorname{ord}_p f$, we call up^r a **p-content** for f, if u is any unit of A. We define the **content** of f to be the product.

$$\prod p^{\operatorname{ord}_p f},$$

the product being taken over all p such that $\operatorname{ord}_p f \neq 0$, or any multiple of this product by a unit of A. Thus the content is well defined up to multiplication by a unit of A. We abbreviate **content** by **cont**.

If $b \in K$, $b \neq 0$, then $\operatorname{cont}(bf) = b \operatorname{cont}(f)$. This is clear. Hence we can write

$$f(X) = c \cdot f_1(X)$$

where $c = \operatorname{cont}(f)$, and $f_1(X)$ has content 1. In particular, all coefficients of f_1 lie in A, and their g.c.d. is 1. We define a polynomial with content 1 to be a **primitive polynomial**.

Theorem 2.1. (Gauss Lemma). *Let A be a factorial ring, and let K be its quotient field. Let $f, g \in K[X]$ be polynomials in one variable. Then*

$$\operatorname{cont}(fg) = \operatorname{cont}(f) \operatorname{cont}(g).$$

Proof. Writing $f = cf_1$ and $g = dg_1$ where $c = \operatorname{cont}(f)$ and $d = \operatorname{cont}(g)$, we see that it suffices to prove: If f, g have content 1, then fg also has content 1, and for this, it suffices to prove that for each prime p, $\operatorname{ord}_p(fg) = 0$. Let

$$f(X) = a_n X^n + \cdots + a_0, \qquad a_n \neq 0,$$

$$g(X) = b_m X^m + \cdots + b_0, \qquad b_m \neq 0,$$

be polynomials of content 1. Let p be a prime of A. It will suffice to prove that p does not divide all coefficients of fg. Let r be the largest integer such that $0 \leqq r \leqq n$, $a_r \neq 0$, and p does not divide a_r. Similarly, let b_s be the coefficient of g farthest to the left, $b_s \neq 0$, such that p does not divide b_s. Consider the coefficient of X^{r+s} in $f(X)g(X)$. This coefficient is equal to

$$c = a_r b_s + a_{r+1} b_{s-1} + \cdots + a_{r-1} b_{s+1} + \cdots$$

and $p \nmid a_r b_s$. However, p divides every other non-zero term in this sum since in each term there will be some coefficient a_i to the left of a_r or some coefficient b_j to the left of b_s. Hence p does not divide c, and our lemma is proved.

We shall now give another proof for the key step in the above argument, namely the statement:

If $f, g \in A[X]$ are primitive (i.e. have content 1) then fg is primitive.

Proof. We have to prove that a given prime p does not divide all the coefficients of fg. Consider reduction mod p, namely the canonical homomorphism $A \to A/(p) = \bar{A}$. Denote the image of a polynomial by a bar, so $f \mapsto \bar{f}$ and $g \mapsto \bar{g}$ under the reduction homomorphism. Then

$$\overline{fg} = \bar{f}\bar{g}.$$

By hypothesis, $\bar{f} \neq 0$ and $\bar{g} \neq 0$. Since $\bar{A}$ is entire, it follows that $\overline{fg} \neq 0$, as was to be shown.

Corollary 2.2. *Let $f(X) \in A[X]$ have a factorization $f(X) = g(X)h(X)$ in $K[X]$. If $c_g = \operatorname{cont}(g)$, $c_h = \operatorname{cont}(h)$, and $g = c_g g_1$, $h = c_h h_1$, then*

$$f(X) = c_g c_h g_1(X) h_1(X),$$

and $c_g c_h$ is an element of A. In particular, if $f, g \in A[X]$ have content 1, then $h \in A[X]$ also.

Proof. The only thing to be proved is $c_g c_h \in A$. But

$$\operatorname{cont}(f) = c_g c_h \operatorname{cont}(g_1 h_1) = c_g c_h,$$

whence our assertion follows.

Theorem 2.3. *Let A be a factorial ring. Then the polynomial ring $A[X]$ in one variable is factorial. Its prime elements are the primes of A and polynomials in $A[X]$ which are irreducible in $K[X]$ and have content 1.*

Proof. Let $f \in A[X]$, $f \neq 0$. Using the unique factorization in $K[X]$ and the preceding corollary, we can find a factorization

$$f(X) = c \cdot p_1(X) \cdots p_r(X)$$

where $c \in A$, and $p_1, \ldots, p_r$ are polynomials in $A[X]$ which are irreducible in $K[X]$. Extracting their contents, we may assume without loss of generality that the content of p_i is 1 for each i. Then $c = \operatorname{cont}(f)$ by the Gauss lemma. This gives us the existence of the factorization. It follows that each $p_i(X)$ is irreducible in $A[X]$. If we have another such factorization, say

$$f(X) = d \cdot q_1(X) \cdots q_s(X),$$

then from the unique factorization in $K[X]$ we conclude that $r = s$, and after a permutation of the factors we have

$$p_i = a_i q_i$$

with elements $a_i \in K$. Since both p_i, q_i are assumed to have content 1, it follows that a_i in fact lies in A and is a unit. This proves our theorem.

Corollary 2.4. *Let A be a factorial ring. Then the ring of polynomials in n variables $A[X_1, \ldots, X_n]$ is factorial. Its units are precisely the units of A, and its prime elements are either primes of A or polynomials which are irreducible in $K[X]$ and have content* 1.

Proof. Induction.

In view of Theorem 2.3, when we deal with polynomials over a factorial ring and having content 1, it is not necessary to specify whether such polynomials are irreducible over A or over the quotient field K. The two notions are equivalent.

Remark 1. The polynomial ring $K[X_1, \ldots, X_n]$ over a field K is not principal when $n \geqq 2$. For instance, the ideal generated by $X_1, \ldots, X_n$ is not principal (trivial proof).

Remark 2. It is usually not too easy to decide when a given polynomial (say in one variable) is irreducible. For instance, the polynomial $X^4 + 4$ is *reducible* over the rational numbers, because

$$X^4 + 4 = (X^2 - 2X + 2)(X^2 + 2X + 2).$$

Later in this book we shall give a precise criterion when a polynomial $X^n - a$ is irreducible. Other criteria are given in the next section.

§3. CRITERIA FOR IRREDUCIBILITY

The first criterion is:

Theorem 3.1. (Eisenstein's Criterion). *Let A be a factorial ring. Let K be its quotient field. Let $f(X) = a_n X^n + \cdots + a_0$ be a polynomial of degree $n \geqq 1$ in $A[X]$. Let p be a prime of A, and assume:*

$$a_n \not\equiv 0 \pmod{p}, \qquad a_i \equiv 0 \pmod{p} \qquad \text{for all} \quad i < n,$$

$$a_0 \not\equiv 0 \pmod{p^2}.$$

Then $f(X)$ is irreducible in $K[X]$.

Proof. Extracting a g.c.d. for the coefficients of f, we may assume without loss of generality that the content of f is 1. If there exists a factorization into factors of degree $\geqq 1$ in $K[X]$, then by the corollary of Gauss' lemma there exists a factorization in $A[X]$, say $f(X) = g(X)h(X)$,

$$g(X) = b_d X^d + \cdots + b_0,$$

$$h(X) = c_m X^m + \cdots + c_0,$$

with $d, m \geqq 1$ and $b_d c_m \neq 0$. Since $b_0 c_0 = a_0$ is divisible by p but not p^2, it follows that one of b_0, c_0 is not divisible by p, say b_0. Then $p | c_0$. Since $c_m b_d = a_n$ is not divisible by p, it follows that p does not divide c_m. Let c_r be the coefficient of h furthest to the right such that $c_r \not\equiv 0 \pmod{p}$. Then

$$a_r = b_0 c_r + b_1 c_{r-1} + \cdots .$$

Since $p \nmid b_0 c_r$ but p divides every other term in this sum, we conclude that $p \nmid a_r$, a contradiction which proves our theorem.

Example. Let a be a non-zero square-free integer $\neq \pm 1$. Then for any integer $n \geqq 1$, the polynomial $X^n - a$ is irreducible over $\mathbf{Q}$. The polynomials $3X^5 - 15$ and $2X^{10} - 21$ are irreducible over $\mathbf{Q}$.

There are some cases in which a polynomial does not satisfy Eisenstein's criterion, but a simple transform of it does.

Example. Let p be a prime number. Then the polynomial

$$f(X) = X^{p-1} + \cdots + 1$$

is irreducible over $\mathbf{Q}$.

Proof. It will suffice to prove that the polynomial $f(X + 1)$ is irreducible over $\mathbf{Q}$. We note that the binomial coefficients

$$\binom{p}{v} = \frac{p!}{v!(p-v)!}, \qquad 1 \leqq v \leqq p - 1,$$

are divisible by p (because the numerator is divisible by p and the denominator is not, and the coefficient is an integer). We have

$$f(X+1) = \frac{(X+1)^p - 1}{(X+1) - 1} = \frac{X^p + pX^{p-1} + \cdots + pX}{X}$$

from which one sees that $f(X + 1)$ satisfies Eisenstein's criterion.

Example. Let E be a field and t an element of some field containing E such that t is transcendental over E. Let K be the quotient field of $E[t]$.

For any integer $n \geqq 1$ the polynomial $X^n - t$ is irreducible in $K[X]$. This comes from the fact that the ring $A = E[t]$ is factorial and that t is a prime in it.

Theorem 3.2. (Reduction Criterion). *Let A, B be entire rings, and let*

$$\varphi: A \to B$$

be a homomorphism. Let K, L be the quotient fields of A and B respectively. Let $f \in A[X]$ be such that $\varphi f \neq 0$ and $\deg \varphi f = \deg f$. If φf is irreducible in $L[X]$, then f does not have a factorization $f(X) = g(X)h(X)$ with

$$g, h \in A[X] \qquad \text{and} \qquad \deg g, \deg h \geqq 1.$$

Proof. Suppose f has such a factorization. Then $\varphi f = (\varphi g)(\varphi h)$. Since $\deg \varphi g \leqq \deg g$ and $\deg \varphi h \leqq \deg h$, our hypothesis implies that we must have equality in these degree relations. Hence from the irreducibility in $L[X]$ we conclude that g or h is an element of A, as desired.

In the preceding criterion, suppose that A is a local ring, i.e. a ring having a unique maximal ideal $\mathfrak{p}$, and that $\mathfrak{p}$ is the kernel of φ. Then from the irreducibility of φf in $L[X]$ we conclude the irreducibility of f in $A[X]$. Indeed, any element of A which does not lie in $\mathfrak{p}$ must be a unit in A, so our last conclusion in the proof can be strengthened to the statement that g or h is a unit in A.

One can also apply the criterion when A is factorial, and in that case deduce the irreducibility of f in $K[X]$.

Example. Let p be a prime number. It will be shown later that $X^p - X - 1$ is irreducible over the field $\mathbf{Z}/p\mathbf{Z}$. Hence $X^p - X - 1$ is irreducible over $\mathbf{Q}$. Similarly,

$$X^5 - 5X^4 - 6X - 1$$

is irreducible over $\mathbf{Q}$.

There is also a routine elementary school test whether a polynomial has a root or not.

Proposition 3.3. (Integral Root Test). *Let A be a factorial ring and K its quotient field. Let*

$$f(X) = a_n X^n + \cdots + a_0 \in A[X].$$

Let $\alpha \in K$ be a root of f, with $\alpha = b/d$ expressed with b, $d \in A$ and b, d relatively prime. Then $b | a_0$ and $d | a_n$. In particular, if the leading coefficient a_n is 1, then a root α must lie in A and divides a_0.

We leave the proof to the reader, who should be used to this one from way back. As an irreducibility test, the test is useful especially for a polynomial of degree 2 or 3, when reducibility is equivalent with the existence of a root in the given field.

§4. HILBERT'S THEOREM

This section proves a basic theorem of Hilbert concerning the ideals of a polynomial ring. We define a commutative ring A to be **Noetherian** if every ideal is finitely generated.

Theorem 4.1. *Let A be a commutative Noetherian ring. Then the polynomial ring $A[X]$ is also Noetherian.*

Proof. Let $\mathfrak{A}$ be an ideal of $A[X]$. Let $\mathfrak{a}_i$ consist of 0 and the set of elements $a \in A$ appearing as leading coefficient in some polynomial

$$a_0 + a_1 X + \cdots + aX^i$$

lying in $\mathfrak{A}$. Then it is clear that $\mathfrak{a}_i$ is an ideal. (If a, b are in $\mathfrak{a}_i$, then $a \pm b$ is in $\mathfrak{a}_i$ as one sees by taking the sum and difference of the corresponding polynomials. If $x \in A$, then $xa \in \mathfrak{a}_i$ as one sees by multiplying the corresponding polynomial by x.) Furthermore we have

$$\mathfrak{a}_0 \subset \mathfrak{a}_1 \subset \mathfrak{a}_2 \subset \cdots,$$

in other words, our sequence of ideals $\{\mathfrak{a}_i\}$ is increasing. Indeed, to see this multiply the above polynomial by X to see that $a \in \mathfrak{a}_{i+1}$.

By criterion (2) of Chapter X, §1, the sequence of ideals $\{\mathfrak{a}_i\}$ stops, say at $\mathfrak{a}_r$:

$$\mathfrak{a}_0 \subset \mathfrak{a}_1 \subset \mathfrak{a}_2 \subset \cdots \subset \mathfrak{a}_r = \mathfrak{a}_{r+1} = \cdots.$$

Let

$$a_{01}, \ldots, a_{0n_0} \quad \text{be generators for} \quad \mathfrak{a}_0,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$a_{r1}, \ldots, a_{rn_r} \quad \text{be generators for} \quad \mathfrak{a}_r.$$

For each $i = 0, \ldots, r$ and $j = 1, \ldots, n_i$ let f_{ij} be a polynomial in $\mathfrak{A}$, of degree i, with leading coefficient a_{ij}. We contend that the polynomials f_{ij} are a set of generators for $\mathfrak{A}$.

Let f be a polynomial of degree d in $\mathfrak{A}$. We shall prove that f is in the ideal generated by the f_{ij}, by induction on d. Say $d \geqq 0$. If $d > r$, then we

note that the leading coefficients of

$$X^{d-r}f_{r1}, \ldots, X^{d-r}f_{rn_r}$$

generate $\mathfrak{a}_d$. Hence there exist elements $c_1, \ldots, c_{n_r} \in A$ such that the polynomial

$$f - c_1X^{d-r}f_{r1} - \cdots - c_{n_r}X^{d-r}f_{rn_r}$$

has degree $< d$, and this polynomial also lies in $\mathfrak{A}$. If $d \leqq r$, we can subtract a linear combination

$$f - c_1f_{d1} - \cdots - c_{n_d}f_{dn_d}$$

to get a polynomial of degree $< d$, also lying in $\mathfrak{A}$. We note that the polynomial we have subtracted from f lies in the ideal generated by the f_{ij}. By induction, we can subtract a polynomial g in the ideal generated by the f_{ij} such that $f - g = 0$, thereby proving our theorem.

We note that if $\varphi: A \to B$ is a surjective homomorphism of commutative rings and A is Noetherian, so is B. Indeed, let $\mathfrak{b}$ be an ideal of B, so $\varphi^{-1}(\mathfrak{b})$ is an ideal of A. Then there is a finite number of generators $(a_1, \ldots, a_n)$ for $\varphi^{-1}(\mathfrak{b})$, and it follows since φ is surjective that $\mathfrak{b} = \varphi(\varphi^{-1}(\mathfrak{b}))$ is generated by $\varphi(a_1), \ldots, \varphi(a_n)$, as desired. As an application, we obtain:

Corollary 4.2. *Let A be a Noetherian commutative ring, and let $B = A[x_1, \ldots, x_m]$ be a commutative ring finitely generated over A. Then B is Noetherian.*

Proof. Use Theorem 4.1 and the preceding remark, representing B as a factor ring of a polynomial ring.

Ideals in polynomial rings will be studied more deeply in Chapter IX. The theory of Noetherian rings and modules will be developed in Chapter X.

§5. PARTIAL FRACTIONS

In this section, we analyze the quotient field of a principal ring, using the factoriality of the ring.

Theorem 5.1. *Let A be a principal entire ring, and let P be a set of representatives for its irreducible elements. Let K be the quotient field of A, and let $\alpha \in K$. For each $p \in P$ there exists an element $\alpha_p \in A$ and an integer $j(p) \geqq 0$, such that $j(p) = 0$ for almost all $p \in P$, α_p and $p^{j(p)}$ are*

relatively prime, and

$$\alpha = \sum_{p \in P} \frac{\alpha_p}{p^{j(p)}}.$$

If we have another such expression

$$\alpha = \sum_{p \in P} \frac{\beta_p}{p^{i(p)}},$$

then $j(p) = i(p)$ *for all* p, *and* $\alpha_p \equiv \beta_p \bmod p^{j(p)}$ *for all* p.

Proof. We first prove existence, in a special case. Let a, b be relatively prime non-zero elements of A. Then there exists x, $y \in A$ such that $xa + yb = 1$. Hence

$$\frac{1}{ab} = \frac{x}{b} + \frac{y}{a}.$$

Hence any fraction c/ab with $c \in A$ can be decomposed into a sum of two fractions (namely cx/b and cy/a) whose denominators divide b and a respectively. By induction, it now follows that any $\alpha \in K$ has an expression as stated in the theorem, except possibly for the fact that p may divide α_p. Canceling the greatest common divisor yields an expression satisfying all the desired conditions.

As for uniqueness, suppose that α has two expressions as stated in the theorem. Let q be a fixed prime in P. Then

$$\frac{\alpha_q}{q^{j(q)}} - \frac{\beta_q}{q^{i(q)}} = \sum_{p \neq q} \frac{\beta_p}{p^{i(p)}} - \frac{\alpha_p}{p^{j(p)}}.$$

If $j(q) = i(q) = 0$, our conditions concerning q are satisfied. Suppose one of $j(q)$ or $i(q) > 0$, say $j(q)$, and say $j(q) \geqq i(q)$. Let d be a least common multiple for all powers $p^{j(p)}$ and $p^{i(p)}$ such that $p \neq q$. Multiply the above equation by $dq^{j(q)}$. We get

$$d(\alpha_q - q^{j(q)-i(q)}\beta_q) = q^{j(q)}\beta$$

for some $\beta \in A$. Furthermore, q does not divide d. If $i(q) < j(q)$ then q divides α_q, which is impossible. Hence $i(q) = j(q)$. We now see that $q^{j(q)}$ divides $\alpha_q - \beta_q$, thereby proving the theorem.

We apply Theorem 5.1 to the polynomial ring $k[X]$ over a field k. We let P be the set of irreducible polynomials, normalized so as to have leading coefficient equal to 1. Then P is a set of representatives for all the irreducible elements of $k[X]$. In the expression given for α in Theorem 5.1, we can now divide α_p by $p^{j(p)}$, i.e. use the Euclidean algorithm, if $\deg \alpha_p \geqq \deg p^{j(p)}$. We denote the quotient field of $k[X]$ by $k(X)$, and call its elements **rational functions**.

Theorem 5.2. *Let $A = k[X]$ be the polynomial ring in one variable over a field k. Let P be the set of irreducible polynomials in $k[X]$ with leading coefficient 1. Then any element f of $k(X)$ has a unique expression*

$$f(X) = \sum_{p \in P} \frac{f_p(X)}{p(X)^{j(p)}} + g(X),$$

where f_p, g are polynomials, $f_p = 0$ if $j(p) = 0$, f_p is relatively prime to p if $j(p) > 0$, and $\deg f_p < \deg p^{j(p)}$ if $j(p) > 0$.

Proof. The existence follows at once from our previous remarks. The uniqueness follows from the fact that if we have two expressions, with elements f_p and φ_p respectively, and polynomials g, h, then $p^{j(p)}$ divides $f_p - \varphi_p$, whence $f_p - \varphi_p = 0$, and therefore $f_p = \varphi_p$, $g = h$.

One can further decompose the term $f_p/p^{j(p)}$ by expanding f_p according to powers of p. One can in fact do something more general.

Theorem 5.3. *Let k be a field and $k[X]$ the polynomial ring in one variable. Let f, $g \in k[X]$, and assume $\deg g \geqq 1$. Then there exist unique polynomials*

$$f_0, f_1, \ldots, f_d \in k[X]$$

such that $\deg f_i < \deg g$ and such that

$$f = f_0 + f_1 g + \cdots + f_d g^d.$$

Proof. We first prove existence. If $\deg g > \deg f$, then we take $f_0 = f$ and $f_i = 0$ for $i > 0$. Suppose $\deg g \leqq \deg f$. We can find polynomials q, r with $\deg r < \deg g$ such that

$$f = qg + r,$$

and since $\deg g \geqq 1$ we have $\deg q < \deg f$. Inductively, there exist polynomials $h_0, h_1, \ldots, h_s$ such that

$$q = h_0 + h_1 g + \cdots + h_s g^s,$$

and hence

$$f = r + h_0 g + \cdots + h_s g^{s+1},$$

thereby proving existence.

As for uniqueness, let

$$f = f_0 + f_1 g + \cdots + f_d g^d = \varphi_0 + \varphi_1 g + \cdots + \varphi_m g^m$$

be two expressions satisfying the conditions of the theorem. Adding terms

equal to 0 to either side, we may assume that $m = d$. Subtracting, we get

$$0 = (f_0 - \varphi_0) + \cdots + (f_d - \varphi_d)g^d.$$

Hence g divides $f_0 - \varphi_0$, and since $\deg(f_0 - \varphi_0) < \deg g$ we see that $f_0 = \varphi_0$. Inductively, take the smallest integer i such that $f_i \neq \varphi_i$ (if such i exists). Dividing the above expression by g^i we find that g divides $f_i - \varphi_i$ and hence that such i cannot exist. This proves uniqueness.

We shall call the expression for f in terms of g in Theorem 5.3 the ***g*-adic expansion** of f. If $g(X) = X$, then the g-adic expansion is the usual expression of f as a polynomial.

Remark. In some sense, Theorem 5.2 redoes what was done in Theorem 8.1 of Chapter I for $\mathbf{Q}/\mathbf{Z}$; that is, express explicitly an element of K/A as a direct sum of its p-components.

§6. SYMMETRIC POLYNOMIALS

Let A be a commutative ring and let $t_1, \ldots, t_n$ be algebraically independent elements over A. Let X be a variable over $A[t_1, \ldots, t_n]$. We form the polynomial

$$\begin{aligned} F(X) &= (X - t_1) \cdots (X - t_n) \\ &= X^n - s_1 X^{n-1} + \cdots + (-1)^n s_n, \end{aligned}$$

where each $s_i = s_i(t_1, \ldots, t_n)$ is a polynomial in $t_1, \ldots, t_n$. Then for instance

$$s_1 = t_1 + \cdots + t_n \qquad \text{and} \qquad s_n = t_1 \cdots t_n.$$

The polynomials $s_1, \ldots, s_n$ are called the **elementary symmetric polynomials** of $t_1, \ldots, t_n$.

We leave it as an easy exercise to verify that s_i **is homogeneous of degree** i **in** $t_1, \ldots, t_n$.

Let σ be a permutation of the integers $(1, \ldots, n)$. Given a polynomial $f(t) \in A[t] = A[t_1, \ldots, t_n]$, we define σf to be

$$\sigma f(t_1, \ldots t_n) = f(t_{\sigma(1)}, \ldots, t_{\sigma(n)}).$$

If σ, τ are two permutations, then $\sigma\tau f = \sigma(\tau f)$ and hence the symmetric group G on n letters operates on the polynomial ring $A[t]$. A polynomial is called **symmetric** if $\sigma f = f$ for all $\sigma \in G$. It is clear that the set of symmetric polynomials is a subring of $A[t]$, which contains the constant polynomials

(i.e. A itself) and also contains the elementary symmetric polynomials $s_1, \ldots, s_n$. We shall see below that $A[s_1, \ldots, s_n]$ is the ring of symmetric polynomials.

Let $X_1, \ldots, X_n$ be variables. We define the **weight** of a monomial

$$X_1^{v_1} \cdots X_n^{v_n}$$

to be $v_1 + 2v_2 + \cdots + nv_n$. We define the weight of a polynomial $g(X_1, \ldots, X_n)$ to be the maximum of the weights of the monomials occurring in g.

Theorem 6.1. *Let $f(t) \in A[t_1, \ldots, t_n]$ be symmetric of degree d. Then there exists a polynomial $g(X_1, \ldots, X_n)$ of weight $\leqq d$ such that*

$$f(t) = g(s_1, \ldots, s_n).$$

If f is homogeneous of degree d, then every monomial occurring in g has weight d.

Proof. By induction on n. The theorem is obvious if $n = 1$, because $s_1 = t_1$. Assume the theorem proved for polynomials in $n - 1$ variables.

If we substitute $t_n = 0$ in the expression for $F(X)$, we find

$$(X - t_1) \cdots (X - t_{n-1})X = X^n - (s_1)_0 X^{n-1} + \cdots + (-1)^{n-1}(s_{n-1})_0 X,$$

where $(s_i)_0$ is the expression obtained by substituting $t_n = 0$ in s_i. We see that $(s_1)_0, \ldots, (s_{n-1})_0$ are precisely the elementary symmetric polynomials in $t_1, \ldots, t_{n-1}$.

We now carry out induction on d. If $d = 0$, our assertion is trivial. Assume $d > 0$, and assume our assertion proved for polynomials of degree $< d$. Let $f(t_1, \ldots, t_n)$ have degree d. There exists a polynomial $g_1(X_1, \ldots, X_{n-1})$ of weight $\leqq d$ such that

$$f(t_1, \ldots, t_{n-1}, 0) = g_1((s_1)_0, \ldots, (s_{n-1})_0).$$

We note that $g_1(s_1, \ldots, s_{n-1})$ has degree $\leqq d$ in $t_1, \ldots, t_n$. The polynomial

$$f_1(t_1, \ldots, t_n) = f(t_1, \ldots, t_n) - g_1(s_1, \ldots, s_{n-1})$$

has degree $\leqq d$ (in $t_1, \ldots, t_n$) and is symmetric. We have

$$f_1(t_1, \ldots, t_{n-1}, 0) = 0.$$

Hence f_1 is divisible by t_n, i.e. contains t_n as a factor. Since f_1 is symmetric, it contains $t_1 \cdots t_n$ as a factor. Hence

$$f_1 = s_n f_2(t_1, \ldots, t_n)$$

for some polynomial f_2, which must be symmetric, and whose degree is

$\leqq d - n < d$. By induction, there exists a polynomial g_2 in n variables and weight $\leqq d - n$ such that

$$f_2(t_1, \ldots, t_n) = g_2(s_1, \ldots, s_n).$$

We obtain

$$f(t) = g_1(s_1, \ldots, s_{n-1}) + s_n g_2(s_1, \ldots, s_n),$$

and each term on the right has weight $\leqq d$. This proves our theorem, except for the last statement which will be left to the reader.

We shall now prove that the elementary symmetric polynomials $s_1, \ldots, s_n$ are algebraically independent over A.

If they are not, take a polynomial $f(X_1, \ldots, X_n) \in A[X]$ of least degree and not equal to 0 such that

$$f(s_1, \ldots, s_n) = 0.$$

Write f as a polynomial in X_n with coefficients in $A[X_1, \ldots, X_{n-1}]$,

$$f(X_1, \ldots, X_n) = f_0(X_1, \ldots, X_{n-1}) + \cdots + f_d(X_1, \ldots, X_{n-1})X_n^d.$$

Then $f_0 \neq 0$. Otherwise, we can write

$$f(X) = X_n \psi(X)$$

with some polynomial ψ, and hence $s_n \psi(s_1, \ldots, s_n) = 0$. From this it follows that $\psi(s_1, \ldots, s_n) = 0$, and ψ has degree smaller than the degree of f.

We substitute s_i for X_i in the above relation, and get

$$0 = f_0(s_1, \ldots, s_{n-1}) + \cdots + f_d(s_1, \ldots, s_{n-1})s_n^d.$$

This is a relation in $A[t_1, \ldots, t_n]$, and we substitute 0 for t_n in this relation. Then all terms become 0 except the first one, which gives

$$0 = f_0((s_1)_0, \ldots, (s_{n-1})_0),$$

using the same notation as in the proof of Theorem 6.1. This is a non-trivial relation between the elementary symmetric polynomials in $t_1, \ldots, t_{n-1}$, a contradiction.

Example. (The Discriminant). Let $f(X) = (X - t_1) \cdots (X - t_n)$. Consider the product

$$\delta(t) = \prod_{i<j} (t_i - t_j).$$

For any permutation σ of $(1, \ldots, n)$ we see at once that

$$\delta^\sigma(t) = \pm \delta(t).$$

Hence $\delta(t)^2$ is symmetric, and we call it the **discriminant**:

$$D_f = D(s_1, \ldots, s_n) = \prod_{i<j} (t_i - t_j)^2.$$

We thus view the discriminant as a polynomial in the elementary symmetric functions. For a continuation of the general theory, see §8. We shall now consider special cases.

Quadratic case. You should verify that for a quadratic polynomial $f(X) = X^2 + bX + c$, one has

$$D = b^2 - 4c.$$

Cubic case. Consider $f(X) = X^3 + aX + b$. We wish to prove that

$$D = -4a^3 - 27b^2.$$

Observe first that D is homogeneous of degree 6 in t_1, t_2. Furthermore, a is homogeneous of degree 2 and b is homogeneous of degree 3. By Theorem 6.1 we know that there exists some polynomial $g(X_2, X_3)$ of weight 6 such that $D = g(a, b)$. The only monomials $X_2^m X_3^n$ of weight 6, i.e. such that $2m + 3n = 6$ with integers $m, n \geqq 0$, are those for which $m = 3$, $n = 0$, or $m = 0$ and $n = 2$. Hence

$$g(X_2, X_3) = vX_2^3 + wX_3^2$$

where v, w are integers which must now be determined.

Observe that the integers v, w are universal, in the sense that for any special polynomial with special values of a, b its discriminant will be given by $g(a, b) = va^3 + wb^2$.

Consider the polynomial

$$f_1(X) = X(X - 1)(X + 1) = X^3 - X.$$

Then $a = -1$, $b = 0$, and $D = va^3 = -v$. But also $D = 4$ by using the definition of the discriminant of the product of the differences of the roots, squared. Hence we get $v = -4$. Next consider the polynomial

$$f_2(X) = X^3 - 1.$$

Then $a = 0$, $b = -1$, and $D = wb^2 = w$. But the three roots of f_2 are the cube roots of unity, namely

$$1, \frac{-1 + \sqrt{-3}}{2}, \frac{-1 - \sqrt{-3}}{2}.$$

Using the definition of the discriminant we find the value $D = -27$. Hence we get $w = -27$. This concludes the proof of the formula for the discriminant of the cubic when there is no X^2 term.

In general, consider a cubic polynomial

$$f(X) = X^3 - s_1X^2 + s_2X - s_3 = (X - t_1)(X - t_2)(X - t_3).$$

We find the value of the discriminant by reducing this case to the simpler case when there is no X^2 term. We make a translation, and let

$$Y = X - \tfrac{1}{3}s_1 \qquad \text{so} \quad X = Y + \tfrac{1}{3}s_1 = Y + \tfrac{1}{3}(t_1 + t_2 + t_3).$$

Then $f(X)$ becomes

$$f(X) = f^*(Y) = Y^3 + aY + b = (Y - u_1)(Y - u_2)(Y - u_3),$$

where $a = u_1u_2 + u_2u_3 + u_1u_3$ and $b = -u_1u_2u_3$, while $u_1 + u_2 + u_3 = 0$. We have

$$u_i = t_i - \tfrac{1}{3}s_1 \qquad \text{for} \quad i = 1, 2, 3,$$

and $u_i - u_j = t_i - t_j$ for all $i \neq j$, so the discriminant is unchanged, and you can easily get the formula in general. Do Exercise 12(b).

§7. MASON-STOTHERS THEOREM AND THE *abc* CONJECTURE

In the early 80s a new trend of thought about polynomials started with the discovery of an entirely new relation. Let $f(t)$ be a polynomial in one variable over the complex numbers if you wish (an algebraically closed field of characteristic 0 would do). We define

$$n_0(f) = \text{number of distinct roots of } f.$$

Thus $n_0(f)$ counts the zeros of f by giving each of them multiplicity 1, and $n_0(f)$ can be small even though $\deg f$ is large.

Theorem 7.1 (Mason-Stothers, [Mas 84], [Sto 81]). *Let $a(t)$, $b(t)$, $c(t)$ be relatively prime polynomials such that $a + b = c$. Then*

$$\max \deg\{a, b, c\} \leqq n_0(abc) - 1.$$

Proof. (Mason) Dividing by c, and letting $f = a/c$, $g = b/c$ we have

$$f + g = 1,$$

where f, g are rational functions. Differentiating we get $f' + g' = 0$, which we rewrite as

$$\frac{f'}{f}f + \frac{g'}{g}g = 0,$$

so that

$$\frac{b}{a} = \frac{g}{f} = -\frac{f'/f}{g'/g}.$$

Let

$$a(t) = c_1 \prod (t - \alpha_i)^{m_i}, \qquad b(t) = c_2 \prod (t - \beta_j)^{n_j}, \qquad c(t) = c_3 \prod (t - \gamma_k)^{r_k}.$$

Then by calculus algebraicized in Exercise 11(c), we get

$$\frac{b}{a} = -\frac{f'/f}{g'/g} = -\frac{\sum \frac{m_i}{t - \alpha_i} - \sum \frac{r_k}{t - \gamma_k}}{\sum \frac{n_j}{t - \beta_j} - \sum \frac{r_k}{t - \gamma_k}}.$$

A common denominator for f'/f and g'/g is given by the product

$$N_0 = \prod (t - \alpha_i) \prod (t - \beta_j) \prod (t - \gamma_k),$$

whose degree is $n_0(abc)$. Observe that $N_0 f'/f$ and $N_0 g'/g$ are both polynomials of degrees at most $n_0(abc) - 1$. From the relation

$$\frac{b}{a} = -\frac{N_0 f'/f}{N_0 g'/g},$$

and the fact that a, b are assumed relatively prime, we deduce the inequality in the theorem.

As an application, let us prove **Fermat's theorem** for polynomials. Thus let $x(t)$, $y(t)$, $z(t)$ be relatively prime polynomials such that one of them has degree $\geqq 1$, and such that

$$x(t)^n + y(t)^n = z(t)^n.$$

We want to prove that $n \leqq 2$. By the Mason-Stothers theorem, we get

$$n \deg x = \deg x(t)^n \leqq \deg x(t) + \deg y(t) + \deg z(t) - 1,$$

and similarly replacing x by y and z on the left-hand side. Adding, we find

$$n(\deg x + \deg y + \deg z) \leqq 3(\deg x + \deg y + \deg z) - 3.$$

This yields a contradiction if $n \geqq 3$.

As another application in the same vein, one has:

Davenport's theorem. *Let f, g be non-constant polynomials such that $f^3 - g^2 \neq 0$. Then*

$$\deg(f^3 - g^2) \geqq \tfrac{1}{2} \deg f - 1.$$

See Exercise 13.

One of the most fruitful analogies in mathematics is that between the integers $\mathbf{Z}$ and the ring of polynomials $F[t]$ over a field F. Evolving from the insights of Mason [Ma 84], Frey [Fr 87], Szpiro, and others, Masser and Oesterle formulated the *abc* conjecture for integers as follows. Let m be a non-zero integer. Define the **radical** of m to be

$$N_0(m) = \prod_{p|m} p,$$

i.e. the product of all the primes dividing m, taken with multiplicity 1.

The *abc* conjecture. *Given $\varepsilon > 0$, there exists a positive number $C(\varepsilon)$ having the following property. For any non-zero relative prime integers a, b, c such that $a + b = c$, we have*

$$\max(|a|, |b|, |c|) \leqq C(\varepsilon) N_0(abc)^{1+\varepsilon}.$$

Observe that the inequality says that many prime factors of a, b, c occur to the first power, and that if "small" primes occur to high powers, then they have to be compensated by "large" primes occurring to the first power. For instance, one might consider the equation

$$2^n \pm 1 = m.$$

For m large, the *abc* conjecture would state that m has to be divisible by large primes to the first power. This phenomenon can be seen in the tables of [BLSTW 83].

Stewart–Tijdeman [ST 86] have shown that it is necessary to have the ε in the formulation of the conjecture. Subsequent examples were communicated to me by Wojtek Jastrzebowski and Dan Spielman as follows.

We have to give examples such that for all $C > 0$ there exist natural numbers a, b, c relatively prime such that $a + b = c$ and $|a| > CN_0(abc)$. But trivially,

$$2^n | (3^{2^n} - 1).$$

We consider the relations $a_n + b_n = c_n$ given by

$$3^{2^n} - 1 = c_n.$$

It is clear that these relations provide the desired examples. Other examples can be constructed similarly, since the role of 3 and 2 can be played by other integers. Replace 2 by some prime, and 3 by an integer $\equiv 1 \bmod p$.

The *abc* conjecture implies what we shall call the

Asymptotic Fermat Theorem. *For all n sufficiently large, the equation*

$$x^n + y^n = z^n$$

has no solution in relatively prime integers $\neq 0$.

The proof follows exactly the same pattern as for polynomials, except that we write things down multiplicatively, and there is a $1 + \varepsilon$ floating around. The extent to which the *abc* conjecture will be proved with an explicit constant $C(\varepsilon)$ (or say $C(1)$ to fix ideas) yields the corresponding explicit determination of the bound for n in the application. We now go into other applications.

Hall's conjecture [Ha 71]. *If u, v are relatively prime non-zero integers such that $u^3 - v^2 \neq 0$, then*

$$|u^3 - v^2| \gg |u|^{1/2-\varepsilon}.$$

The symbol $\gg$ means that the left-hand side is $\geqq$ the right-hand side times a constant depending only on ε. Again the proof is immediate from the *abc* conjecture. Actually, the hypothesis that u, v are relatively prime is not necessary; the general case can be reduced to the relatively prime case by extracting common factors, and Hall stated his conjecture in this more general way. However, he also stated it without the epsilon in the exponent, and that does not work, as was realized later. As in the polynomial case, Hall's conjecture describes how small $|u^3 - v^2|$ can be, and the answer is not too small, as described by the right-hand side.

The Hall conjecture can also be interpreted as giving a bound for integral relatively prime solutions of

$$v^2 = u^3 + b \quad \text{with integral } b.$$

Then we find

$$|u| \ll |b|^{2+\varepsilon}.$$

More generally, in line with conjectured inequalities from Lang–Waldschmidt [La 78], let us fix non-zero integers A, B and let u, v, k, m, n be variable, with u, v relatively prime and $mv > m + n$. Put

$$Au^m + Bv^n = k.$$

By the *abc* conjecture, one derives easily that

$$(1) \qquad |u| \ll N_0(k)^{\frac{n}{mn-(m+n)}(1+\varepsilon)} \qquad \text{and} \qquad |v| \ll N_0(k)^{\frac{m}{mn-(m+n)}(1+\varepsilon)}.$$

From this one gets

$$|k| \ll N_0(k)^{\frac{mn}{mn-(m+n)}(1+\varepsilon)}.$$

The Hall conjecture is a special case after we replace $N_0(k)$ with $|k|$, because $N_0(k) \leqq |k|$.

Next take $m = 3$ and $n = 2$, but take $A = 4$ and $B = -27$. In this case we write

$$D = 4u^3 - 27v^2$$

and we get

$$|u| \ll N_0(D)^{2+\varepsilon} \quad \text{and} \quad |v| \ll N_0(D)^{3+\varepsilon}. \tag{2}$$

These inequalities are supposed to hold at first for u, v relatively prime. Suppose we allow u, v to have some bounded common factor, say d. Write

$$u = u'd \quad \text{and} \quad v = v'd$$

with u', v' relatively prime. Then

$$D = 4d^3u'^3 - 27d^2v'^2.$$

Now we can apply inequality (1) with $A = 4d^3$ and $B = -27d^2$, and we find the same inequalities (2), with the constant implicit in the sign $\ll$ depending also on d, or on some fixed bound for such a common factor. Under these circumstances, we call inequalities (2) the **generalized Szpiro conjecture**.

The original Szpiro conjecture was stated in a more sophisticated situation, cf. [La 90] for an exposition, and Szpiro's inequality was stated in the form

$$|D| \ll N(D)^{6+\varepsilon},$$

where $N(D)$ is a more subtle invariant, but for our purposes, it is sufficient and much easier to use the radical $N_0(D)$.

The point of D is that it occurs as a discriminant. The trend of thoughts in the direction we are discussing was started by Frey [Fr 87], who associated with each solution of $a + b = c$ the polynomial

$$x(x - a)(x + b),$$

which we call the **Frey polynomial**. (Actually Frey associated the curve defined by the equation $y^2 = x(x - a)(x + b)$, for much deeper reasons, but only the polynomial on the right-hand side will be needed here.) The discriminant of the polynomial is the product of the differences of the roots squared, and so

$$D = (abc)^2.$$

We make a translation

$$\xi = x + \frac{b - a}{3}$$

to get rid of the x^2-term, so that our polynomial can be rewritten

$$\xi^3 - \gamma_2\xi - \gamma_3,$$

where γ_2, γ_3 are homogeneous in a, b of appropriate weight. The discriminant does not change because the roots of the polynomial in ξ are

translations of the roots of the polynomial in x. Then

$$D = 4\gamma_2^3 - 27\gamma_3^2.$$

The translation with $(b - a)/3$ introduces a small denominator. One may avoid this denominator by using the polynomial $x(x - 3a)(x - 3b)$, so that γ_2, γ_3 then come out to be integers, and one can apply the generalized Szpiro conjecture to the discriminant, which then has an extra factor $D = 3^6(abc)^2$.

It is immediately seen that the generalized Szpiro conjecture implies asymptotic Fermat. Conversely:

Generalized Szpiro implies the abc conjecture.

Indeed, the correspondence $(a, b) \leftrightarrow (\gamma_2, \gamma_3)$ is invertible, and has the "right" weight. A simple algebraic manipulation shows that the generalized Szpiro estimates on γ_2, γ_3 imply the desired estimates on $|a|$, $|b|$. (Do Exercise 14.) From the equivalence between *abc* and generalized Szpiro, one can use the examples given earlier to show that the epsilon is needed in the Szpiro conjecture.

Finally, note that the polynomial case of the Mason-Stothers theorem and the case of integers are not independent, or specifically the Davenport theorem and Hall's conjecture are related. Examples in the polynomial case parametrize cases with integers when we substitute integers for the variables. Such examples are given in [BCHS 65], one of them (due to Birch) being

$$f(t) = t^6 + 4t^4 + 10t^2 + 6 \qquad \text{and} \qquad g(t) = t^9 + 6t^7 + 21t^5 + 35t^3 + \tfrac{63}{2}t,$$

whence

$$\deg(f(t)^3 - g(t)^2) = \tfrac{1}{2}\deg f + 1.$$

This example shows that Davenport's inequality is best possible, because the degree attains the lowest possible value permissible under the theorem. Substituting large integral values of $t \equiv 2 \bmod 4$ gives examples of similarly low values for $x^3 - y^2$. For other connections of all these matters, cf. [La 90].

Bibliography

[BCHS 65] B. Birch, S. Chowla, M. Hall, and A. Schinzel, On the difference $x^3 - y^2$, *Norske Vid. Selsk. Forrh.* **38** (1965) pp. 65–69

[BLSTW 83] J. Brillhart, D. H. Lehmer, J. L. Selfridge, B. Tuckerman, and S. Wagstaff Jr., Factorization of $b^n \pm 1$, $b = 2, 3, 5, 6, 7, 10, 11$ up to high powers, *Contemporary Mathematics* Vol. **22**, AMS, Providence, RI, 1983

[Dav 65] H. Davenport, On $f^3(t) - g^2(t)$, *Norske Vid. Selsk. Forrh.* **38** (1965) pp. 86–87

[Fr 87] G. Frey, Links between solutions of $A - B = C$ and elliptic curves, *Number Theory, Lecture Notes* **1380**, Springer-Verlag, New York, 1989 pp. 31–62

[Ha 71] M. HALL, The diophantine equation $x^3 - y^2 = k$, *Computers and Number Theory*, ed. by A. O. L. Atkin and B. Birch, Academic Press, London 1971 pp. 173–198

[La 90] S. LANG, Old and new conjectured diophantine inequalities, *Bull. AMS* Vol. **23** No. 1 (1990) pp. 37–75

[Ma 84a] R. C. MASON, Equations over function fields, Springer Lecture Notes **1068** (1984), pp. 149–157; in *Number Theory, Proceedings of the Noordwijkerhout*, 1983

[Ma 84b] R. C. MASON, Diophantine equations over function fields, *London Math. Soc. Lecture Note Series* Vol. **96**, Cambridge University Press, Cambridge, 1984

[Ma 84c] R. C. MASON, The hyperelliptic equation over function fields, *Math. Proc. Cambridge Philos. Soc.* **93** (1983) pp. 219–230

[Si 88] J. SILVERMAN, Wieferich's criterion and the *abc* conjecture, *Journal of Number Theory* **30** (1988) pp. 226–237

[ST 86] C. L. STEWART and R. TIJDEMAN, On the Oesterle–Masser Conjecture, *Mon. Math.* **102** (1986) pp. 251–257

See additional references at the end of the chapter.

§8. THE RESULTANT

In this section, we assume that the reader is familiar with determinants. The theory of determinants will be covered later. The section can be viewed as giving further examples of symmetric functions.

Let A be a commutative ring and let $v_0, \ldots, v_n$, $w_0, \ldots, w_m$ be algebraically independent over A. We form two polynomials:

$$f_v(X) = v_0 X^n + \cdots + v_n,$$

$$g_w(X) = w_0 X^m + \cdots + w_m.$$

We define the **resultant** of (v, w), or of f_v, g_w, to be the determinant

$$\begin{vmatrix}
v_0 & v_1 & \cdots & v_n & & \\
 & v_0 & v_1 & \cdots & v_n & \\
 & & \cdots & \cdots & \cdots & \cdots \\
 & & & v_0 & v_1 & \cdots & v_n \\
w_0 & w_1 & \cdots & w_m & & \\
 & w_0 & w_1 & \cdots & w_m & \\
 & & \cdots & \cdots & \cdots & \cdots \\
 & & & w_0 & w_1 & \cdots & w_m
\end{vmatrix}$$

(m rows of the v-coefficients, n rows of the w-coefficients; $m + n$ columns)

The blank spaces are supposed to be filled with zeros.

If we substitute elements $(a) = (a_0, \ldots, a_n)$ and $(b) = (b_0, \ldots, b_m)$ in A for (v) and (w) respectively in the coefficients of f_v and g_w, then we obtain polynomials f_a and g_b with coefficients in A, and *we define their* **resultant** *to be the determinant obtained by substituting* (a) *for* (v) *and* (b) *for* (w) *in the determinant.* We shall write the resultant of f_v, g_w in the form

$$\operatorname{Res}(f_v, g_w) \quad \text{or} \quad R(v, w).$$

The resultant $\operatorname{Res}(f_a, g_b)$ is then obtained by substitution of (a), (b) for (v), (w) respectively.

We observe that $R(v, w)$ is a polynomial with integer coefficients, i.e. we may take $A = \mathbf{Z}$. If z is a variable, then

$$R(zv, w) = z^m R(v, w) \qquad \text{and} \qquad R(v, zw) = z^n R(v, w)$$

as one sees immediately by factoring out z from the first m rows (resp. the last n rows) in the determinant. Thus R is homogeneous of degree m in its first set of variables, and homogeneous of degree n in its second set of variables. Furthermore, $R(v, w)$ contains the monomial

$$v_0^m w_m^n$$

with coefficient 1, when expressed as a sum of monomials.

If we substitute 0 for v_0 and w_0 in the resultant, we obtain 0, because the first column of the determinant vanishes.

Let us work over the integers $\mathbf{Z}$. We consider the linear equations

$$\begin{aligned}
X^{m-1}f_v(X) &= v_0X^{n+m-1} + v_1X^{n+m-2} + \cdots + v_nX^{m-1} \\
X^{m-2}f_v(X) &= \qquad v_0X^{n+m-2} + \cdots + v_nX^{m-2} \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
f_v(X) &= \qquad\qquad v_0X^n + \cdots + v_n \\
X^{n-1}g_w(X) &= w_0X^{n+m-1} + w_1X^{n+m-2} + \cdots + w_mX^{n-1} \\
X^{n-2}g_w(X) &= \qquad w_0X^{n+m-2} + \cdots + w_mX^{n-2} \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
g_w(X) &= \qquad\qquad w_0X^m + \cdots + w_m.
\end{aligned}$$

Let C be the column vector on the left-hand side, and let

$$C_0, \ldots, C_{m+n}$$

be the column vectors of coefficients. Our equations can be written

$$C = X^{n+m-1}C_0 + \cdots + 1 \cdot C_{m+n}.$$

By Cramer's rule, applied to the last coefficient which is $= 1$,

$$R(v, w) = \det(C_0, \ldots, C_{m+n}) = \det(C_0, \ldots, C_{m+n-1}, C).$$

From this we see that there exist polynomials $\varphi_{v,w}$ *and* $\psi_{v,w}$ *in* $\mathbf{Z}[v, w][X]$ *such that*

$$\varphi_{,v,w}f_v + \psi_{v,w}g_w = R(v,w) = \operatorname{Res}(f_v, f_w).$$

Note that $R(v, w) \in \mathbf{Z}[v, w]$ but that the polynomials on the left-hand side involve the variable X.

If $\lambda\colon \mathbf{Z}[v, w] \to A$ is a homomorphism into a commutative ring A and we let $\lambda(v) = (a)$, $\lambda(w) = (b)$, then

$$\varphi_{a,b}f_a + \psi_{a,b}g_b = R(a, b) = \operatorname{Res}(f_a, f_b).$$

Thus from the universal relation of the resultant over $\mathbf{Z}$ we obtain a similar relation for every pair of polynomials, in any commutative ring A.

Proposition 8.1. *Let* K *be a subfield of a field* L, *and let* f_a, g_b *be polynomials in* $K[X]$ *having a common root* ξ *in* L. *Then* $R(a, b) = 0$.

Proof. If $f_a(\xi) = g_b(\xi) = 0$, then we substitute ξ for X in the expression obtained for $R(a, b)$ and find $R(a, b) = 0$.

Next, we shall investigate the relationship between the resultant and the roots of our polynomials f_v, g_w. We need a lemma.

Lemma 8.2. *Let* $h(X_1, \ldots, X_n)$ *be a polynomial in* n *variables over the integers* $\mathbf{Z}$. *If* h *has the value* 0 *when we substitute* X_1 *for* X_2 *and leave the other* X_i *fixed* $(i \neq 2)$, *then* $h(X_1, \ldots, X_n)$ *is divisible by* $X_1 - X_2$ *in* $\mathbf{Z}[X_1, \ldots, X_n]$.

Proof. Exercise for the reader.

Let $v_0, t_1, \ldots, t_n, w_0, u_1, \ldots, u_m$ be algebraically independent over $\mathbf{Z}$ and form the polynomials

$$f_v = v_0(X - t_1) \cdots (X - t_n) = v_0X^n + \cdots + v_n,$$
$$g_w = w_0(X - u_1) \cdots (X - u_m) = w_0X^m + \cdots + w_m.$$

Thus we let

$$v_i = (-1)^i v_0 s_i(t) \qquad \text{and} \qquad w_j = (-1)^j w_0 s_j(u).$$

We leave to the reader the easy verification that

$$v_0, v_1, \ldots, v_n, w_0, w_1, \ldots, w_m$$

are algebraically independent over $\mathbf{Z}$.

Proposition 8.3. *Notation being as above, we have*

$$\operatorname{Res}(f_v, g_w) = v_0^m w_0^n \prod_{i=1}^{n} \prod_{j=1}^{m} (t_i - u_j).$$

Proof. Let S be the expression on the right-hand side of the equality in the statement of the proposition.

Since $R(v, w)$ is homogeneous of degree m in its first variables, and homogeneous of degree n in its second variables, it follows that

$$R = v_0^m w_0^n h(t, u)$$

where $h(t, u) \in \mathbf{Z}[t, u]$. By Proposition 8.1, the resultant vanishes when we substitute t_i for u_j ($i = 1, \ldots, n$ and $j = 1, \ldots, m$), whence by the lemma, viewing R as an element of $\mathbf{Z}[v_0, w_0, t, u]$ it follows that R is divisible by $t_i - u_j$ for each pair (i, j). Hence S divides R in $\mathbf{Z}[v_0, w_0, t, u]$, because $t_i - u_j$ is obviously a prime in that ring, and different pairs (i, j) give rise to different primes.

From the product expression for S, namely

$$S = v_0^m w_0^n \prod_{i=1}^{n} \prod_{j=1}^{m} (t_i - u_j), \tag{1}$$

we obtain

$$\prod_{i=1}^{n} g(t_i) = w_0^n \prod_{i=1}^{n} \prod_{j=1}^{m} (t_i - u_j),$$

whence

$$S = v_0^m \prod_{i=1}^{n} g(t_i). \tag{2}$$

Similarly,

$$S = (-1)^{nm} w_0^n \prod_{j=1}^{m} f(u_j). \tag{3}$$

From (2) we see that S is homogeneous and of degree n in (w), and from (3) we see that S is homogeneous and of degree m in (v). Since R has exactly the same homogeneity properties, and is divisible by S, it follows that $R = cS$ for some integer c. Since both R and S have a monomial $v_0^m w_m^n$ occurring in them with coefficient 1, it follows that $c = 1$, and our proposition is proved.

We also note that the three expressions found for S above now give us a factorization of R. We also get a converse for Proposition 8.1.

Corollary 8.4. *Let f_a, g_b be polynomials with coefficients in a field K, such that $a_0 b_0 \neq 0$, and such that f_a, g_b split in factors of degree 1 in $K[X]$. Then* $\operatorname{Res}(f_a, g_b) = 0$ *if and only if f_a and g_b have a root in common.*

Proof. Assume that the resultant is 0. If

$$f_a = a_0(X - \alpha_1) \cdots (X - \alpha_n),$$

$$g_b = b_0(X - \beta_1) \cdots (X - \beta_n),$$

is the factorization of f_a, g_b, then we have a homomorphism

$$\mathbf{Z}[v_0, t, w_0, u] \to K$$

such that $v_0 \mapsto a_0$, $w_0 \mapsto b_0$, $t_i \mapsto \alpha_i$, and $u_j \mapsto \beta_j$ for all i, j. Then

$$0 = \operatorname{Res}(f_a, g_b) = a_0^m b_0^n \prod_i \prod_j (\alpha_i - \beta_j),$$

whence f_a, f_b have a root in common. The converse has already been proved.

We deduce one more relation for the resultant in a special case. Let f_v be as above,

$$f_v(X) = v_0 X^n + \cdots + v_n = v_0(X - t_1) \cdots (X - t_n).$$

From (2) we know that if f_v' is the derivative of f_v, then

$$\operatorname{Res}(f_v, f_v') = v_0^{n-1} \prod_i f'(t_i). \tag{4}$$

Using the product rule for differentiation, we find:

$$f_v'(X) = \sum_i v_0(X - t_1) \cdots \widehat{(X - t_i)} \cdots (X - t_n),$$

$$f_v'(t_i) = v_0(t_i - t_1) \cdots \widehat{(t_i - t_i)} \cdots (t_i - t_n),$$

where a roof over a term means that this term is to be omitted.

We define the **discriminant** of f_v to be

$$D(f_v) = D(v) = (-1)^{n(n-1)/2} v_0^{2n-2} \prod_{i \neq j} (t_i - t_j).$$

Proposition 8.5. *Let f_v be as above and have algebraically independent coefficients over $\mathbf{Z}$. Then*

$$\operatorname{Res}(f_v, f_v') = v_0^{2n-1} \prod_{i \neq j} (t_i - t_j) = (-1)^{n(n-1)/2} v_0 D(f_v). \tag{5}$$

Proof. One substitutes the expression obtained for $f_v'(t_i)$ into the product (4). The result follows at once.

When we substitute 1 for v_0, we find that the discriminant as we defined it in the preceding section coincides with the present definition. In particular, we find an explicit formula for the discriminant. The formulas in the special case of polynomials of degree 2 and 3 will be given as exercises.

Note that the discriminant can also be written as the product

$$D(f_v) = v_0^{2n-2} \prod_{i<j} (t_i - t_j)^2.$$

Serre once pointed out to me that the sign $(-1)^{n(n-1)/2}$ was missing in the first edition of this book, and that this sign error is quite common in the literature, occurring as it does in van der Waerden, Samuel, and Hilbert (but not in his collected works, corrected by Olga Taussky); on the other hand the sign is correctly given in Weber's *Algebra*, Vol. I, 50.

For a continuation of this section, see Chapter IX, §3 and §4.

§9. POWER SERIES

Let X be a letter, and let G be the monoid of functions from the set $\{X\}$ to the natural numbers. If $v \in \mathbf{N}$, we denote by X^v the function whose value at X is v. Then G is a multiplicative monoid, already encountered when we discussed polynomials. Its elements are $X^0, X^1, X^2, \ldots, X^v, \ldots$.

Let A be a commutative ring, and let $A[[X]]$ be the set of functions from G into A, without any restriction. Then an element of $A[[X]]$ may be viewed as assigning to each monomial X^v a coefficient $a_v \in A$. We denote this element by

$$\sum_{v=0}^{\infty} a_v X^v.$$

The summation symbol is not a sum, of course, but we shall write the above expression also in the form

$$a_0 X^0 + a_1 X^1 + \cdots$$

and we call it a **formal power series** with coefficients in A, in one variable. We call $a_0, a_1, \ldots$ its coefficients.

Given two elements of $A[[X]]$, say

$$\sum_{v=0}^{\infty} a_v X^v \quad \text{and} \quad \sum_{\mu=0}^{\infty} b_\mu X^\mu,$$

we define their product to be

$$\sum_{i=0}^{\infty} c_i X^i$$

where

$$c_i = \sum_{v+\mu=i} a_v b_\mu.$$

Just as with polynomials, one defines their sum to be

$$\sum_{v=0}^{\infty} (a_v + b_v) X^v.$$

Then we see that the power series form a ring, the proof being the same as for polynomials.

One can also construct the power series ring in several variables $A[[X_1, \ldots, X_n]]$ in which every element can be expressed in the form

$$\sum_{(v)} a_{(v)} X_1^{v_1} \cdots X_n^{v_n} = \sum a_{(v)} M_{(v)}(X_1, \ldots, X_n)$$

with unrestricted coefficients $a_{(v)}$ in bijection with the n-tuples of integers $(v_1, \ldots, v_n)$ such that $v_i \geqq 0$ for all i. It is then easy to show that there is an isomorphism between $A[[X_1, \ldots, X_n]]$ and the repeated power series ring $A[[X_1]] \cdots [[X_n]]$. We leave this as an exercise for the reader.

The next theorem will give an analogue of the Euclidean algorithm for power series. However, instead of dealing with power series over a field, it is important to have somewhat more general coefficients for certain applications, so we have to introduce a little more terminology.

Let A be a ring and I an ideal. We assume that

$$\bigcap_{v=1}^{\infty} I^v = \{0\}.$$

We can view the powers I^v as defining neighborhoods of 0 in A, and we can transpose the usual definition of Cauchy sequence in analysis to this situation, namely: we define a sequence $\{a_n\}$ in A to be **Cauchy** if given some power I^v there exists an integer N such that for all $m, n \geqq N$ we have

$$a_m - a_n \in I^v.$$

Thus I^v corresponds to the given ϵ of analysis. Then we have the usual notion of **convergence** of a sequence to an element of A. One says that A is **complete in the I-adic topology** if every Cauchy sequence converges.

Perhaps the most important example of this situation is when A is a local ring and $I = \mathfrak{m}$ is its maximal ideal. By a **complete local ring**, one always means a local ring which is complete in the $\mathfrak{m}$-adic topology.

Let k be a field. Then the power series ring

$$R = k[[X_1, \ldots, X_n]]$$

in n variables is such a complete local ring. Indeed, let $\mathfrak{m}$ be the ideal generated by the variables $X_1, \ldots, X_n$. Then $R/\mathfrak{m}$ is naturally isomorphic to the field k itself, so $\mathfrak{m}$ is a maximal ideal. Furthermore, any power series of the form

$$f(X) = c_0 - f_1(X)$$

with $c_0 \in k$, $c_0 \neq 0$ and $f_1(X) \in \mathfrak{m}$ is invertible. To prove this, one may first assume without loss of generality that $c_0 = 1$. Then

$$(1 - f_1(X))^{-1} = 1 + f_1(X) + f_1(X)^2 + f_1(X)^3 + \cdots$$

gives the inverse. Thus we see that $\mathfrak{m}$ is the unique maximal ideal and R is local. It is immediately verified that R is complete in the sense we have just defined. The same argument shows that if k is not a field but c_0 is invertible in k, then again $f(X)$ is invertible.

Again let A be a ring. We may view the power series ring in n variables ($n > 1$) as the ring of power series in one variable X_n over the ring of power series in $n - 1$ variables, that is we have a natural identification

$$A[[X_1, \ldots, X_n]] = A[[X_1, \ldots, X_{n-1}]][[X_n]].$$

If $A = k$ is a field, the ring $k[[X_1, \ldots, X_{n-1}]]$ is then a complete local ring. More generally, if $\mathfrak{o}$ is a complete local ring, then the power series ring $\mathfrak{o}[[X]]$ is a complete local ring, whose maximal ideal is $(\mathfrak{m}, X)$ where $\mathfrak{m}$ is the maximal ideal of $\mathfrak{o}$. Indeed, if a power series $\sum a_v X^v$ has unit constant

term $a_0 \in \mathfrak{o}^*$, then the power series is a unit in $\mathfrak{o}[[X]]$, because first, without loss of generality, we may assume that $a_0 = 1$, and then we may invert $1 + h$ with $h \in (\mathfrak{m}, X)$ by the geometric series $1 - h + h^2 - h^3 + \cdots$.

In a number of problems, it is useful to reduce certain questions about power series in several variables over a field to questions about power series in one variable over the more complicated ring as above. We shall now apply this decomposition to the Euclidean algorithm for power series.

Theorem 9.1. *Let $\mathfrak{o}$ be a complete local ring with maximal ideal $\mathfrak{m}$. Let*

$$f(X) = \sum_{i=0}^{\infty} a_i X^i$$

be a power series in $\mathfrak{o}[[X]]$ (one variable), such that not all a_i lie in $\mathfrak{m}$. Say $a_0, \ldots, a_{n-1} \in \mathfrak{m}$, and $a_n \in \mathfrak{o}^*$ is a unit. Given $g \in \mathfrak{o}[[X]]$ we can solve the equation

$$g = qf + r$$

uniquely with $q \in \mathfrak{o}[[X]]$, $r \in \mathfrak{o}[X]$, and $\deg r \leqq n - 1$.

Proof (Manin). Let α and τ be the projections on the beginning and tail end of the series, given by

$$\alpha: \sum b_i X^i \mapsto \sum_{i=0}^{n-1} b_i X^i = b_0 + b_1 X + \cdots + b_{n-1} X^{n-1},$$

$$\tau: \sum b_i X^i \mapsto \sum_{i=n}^{\infty} b_i X^{i-n} = b_n + b_{n+1} X + b_{n+2} X^2 + \cdots.$$

Note that $\tau(hX^n) = h$ for any $h \in \mathfrak{o}[[X]]$; and h is a polynomial of degree $< n$ if and only if $\tau(h) = 0$.

The existence of q, r is equivalent with the condition that there exists q such that

$$\tau(g) = \tau(qf).$$

Hence our problem is equivalent with solving

$$\tau(g) = \tau\big(q\alpha(f)\big) + \tau\big(q\tau(f)X^n\big) = \tau\big(q\alpha(f)\big) + q\tau(f).$$

Note that $\tau(f)$ is invertible. Put $Z = q\tau(f)$. Then the above equation is equivalent with

$$\tau(g) = \tau\left(Z\frac{\alpha(f)}{\tau(f)}\right) + Z = \left(I + \tau \circ \frac{\alpha(f)}{\tau(f)}\right)Z.$$

Note that

$$\tau \circ \frac{\alpha(f)}{\tau(f)}: \mathfrak{o}[[X]] \to \mathfrak{m}\mathfrak{o}[[X]],$$

because $\alpha(f)/\tau(f) \in \mathfrak{m}\mathfrak{o}[[X]]$. We can therefore invert to find Z, namely

$$Z = \left(I + \tau \circ \frac{\alpha(f)}{\tau(f)}\right)^{-1} \tau(g),$$

which proves both existence and uniqueness and concludes the proof.

Theorem 9.2. (Weierstrass Preparation). *The power series f in the previous theorem can be written uniquely in the form*

$$f(X) = (X^n + b_{n-1}X^{n-1} + \cdots + b_0)u,$$

where $b_i \in \mathfrak{m}$, and u is a unit in $\mathfrak{o}[[X]]$.

Proof. Write uniquely

$$X^n = qf + r,$$

by the Euclidean algorithm. Then q is invertible, because

$$q = c_0 + c_1 X + \cdots,$$

$$f = \cdots + a_n X^n + \cdots,$$

so that

$$1 \equiv c_0 a_n \pmod{\mathfrak{m}},$$

and therefore c_0 is a unit in $\mathfrak{o}$. We obtain $qf = X^n - r$, and

$$f = q^{-1}(X^n - r),$$

with $r \equiv 0 \pmod{\mathfrak{m}}$. This proves the existence. Uniqueness is immediate.

The integer n in Theorems 9.1 and 9.2 is called the **Weierstrass degree** of f, and is denoted by $\deg_W f$. We see that a power series not all of whose coefficients lie in $\mathfrak{m}$ can be expressed as a product of a polynomial having the given Weierstrass degree, times a unit in the power series ring. Furthermore, all the coefficients of the polynomial except the leading one lie in the maximal ideal. Such a polynomial is called **distinguished**, or a **Weierstrass polynomial.**

Remark. I rather like the use of the Euclidean algorithm in the proof of the Weierstrass Preparation theorem. However, one can also give a direct proof exhibiting explicitly the recursion relations which solve for the coefficients of u, as follows. Write $u = \sum c_i X^i$. Then we have to solve the equations

$$b_0 c_0 = a_0,$$

$$b_0 c_1 + b_1 c_0 = a_1,$$

$$\cdots$$

$$b_0 c_{n-1} + \cdots + b_{n-1} c_0 = a_{n-1},$$

$$b_0 c_n + \cdots + c_0 = a_n,$$

$$b_0 c_{n+1} + \cdots + c_1 = a_{n+1},$$

$$\cdots$$

In fact, the system of equations has a unique solution mod $\mathfrak{m}^r$ for each positive integer r, after selecting c_0 to be a unit, say $c_0 = 1$. Indeed, from the first n equations (from 0 to $n-1$) we see that $b_0, \ldots, b_{n-1}$ are uniquely determined to be 0 mod $\mathfrak{m}$. Then $c_n, c_{n+1}, \ldots$ are uniquely determined mod $\mathfrak{m}$ by the subsequent equations. Now inductively, suppose we have shown that the coefficients b_i, c_j are uniquely determined mod $\mathfrak{m}^r$. Then one sees immediately that from the conditions $a_0, \ldots, a_{n-1} \equiv 0 \bmod \mathfrak{m}$ the first n equations define b_i uniquely mod $\mathfrak{m}^{r+1}$ because all $b_i \equiv 0 \bmod \mathfrak{m}$. Then the subsequent equations define $c_j \bmod \mathfrak{m}^{r+1}$ uniquely from the values of $b_i \bmod \mathfrak{m}^{r+1}$ and $c_j \bmod \mathfrak{m}^r$. The unique system of solutions mod $\mathfrak{m}^r$ for each r then defines a solution in the projective limit, which is the complete local ring.

We now have all the tools to deal with unique factorization in one important case.

Theorem 9.3. *Let k be a field. Then $k[[X_1, \ldots, X_n]]$ is factorial.*

Proof. Let $f(x) = f(X_1, \ldots, X_n) \in k[[X]]$ be $\neq 0$. After making a sufficiently general linear change of variables (when k is infinite)

$$x_i = \sum c_{ij} Y_j \quad \text{with} \quad c_{ij} \in k,$$

we may assume without loss of generality that $f(0, \ldots, 0, x_n) \neq 0$. (When k is finite, one has to make a non-linear change, cf. Theorem 2.1 of Chapter VIII.) Indeed, if we write $f(X) = f_d(X) +$ higher terms, where $f_d(X)$ is a homogeneous polynomial of degree $d \geqq 0$, then changing the variables as above preserves the degree of each homogeneous component of f, and since k is assumed infinite, the coefficients c_{ij} can be taken so that in fact each power Y_i^d $(i = 1, \ldots, n)$ occurs with non-zero coefficient.

We now proceed by induction on n. Let $R_n = k[[X_1, \ldots, X_n]]$ be the power series in n variables, and assume by induction that R_{n-1} is factorial. By Theorem 9.2, write $f = gu$ where u is a unit and g is a Weierstrass polynomial in $R_{n-1}[X_n]$. By Theorem 2.3, $R_{n-1}[X_n]$ is factorial, and so we can write g as a product of irreducible elements $g_1, \ldots, g_r \in R_{n-1}[X_n]$, so $f = g_1 \cdots g_r u$, where the factors g_i are uniquely determined up to multiplication by units. This proves the existence of a factorization. As to uniqueness, suppose f is expressed as a product of irreducible elements in R_n, $f = f_1 \cdots f_s$. Then $f_q(0, \ldots, 0, x_n) \neq 0$ for each $q = 1, \ldots, s$, so we can write $f_q = h_q u_q'$ where u_q' is a unit and h_q is a Weierstrass polynomial, necessarily irreducible in $R_{n-1}[X_n]$. Then $f = gu = \prod h_q \prod u_q'$ with g and all h_q Weierstrass polynomials. By Theorem 9.2, we must have $g = \prod h_q$, and since $R_{n-1}[X_n]$ is factorial, it follows that the polynomials h_q are the same as the polynomials g_i, up to units. This proves uniqueness.

Remark. As was pointed out to me by Dan Anderson, I incorrectly stated in a previous printing that if $\mathfrak{O}$ is a factorial complete local ring, then $\mathfrak{O}[[X]]$ is also factorial. This assertion is false, as shown by the example

$$k(t)[[X_1, X_2, X_3]]/(X_1^2 + X_2^2 + X_3^2)$$

due to P. Salmon, *Su un problema posto da P. Samuel*, Atti Acad. Naz. Lincei Rend. Cl. Sc. Fis. Matem. **40(8)** (1966) pp. 801–803. It is true that if $\mathfrak{O}$ is a regular local ring *in addition* to being complete, then $\mathfrak{O}[[X]]$ is factorial, but this is a deeper theorem. The simple proof I gave for the power series over a field is classical. I chose the exposition in [GrH 78].

Theorem 9.4. *If A is Noetherian, then $A[[X]]$ is also Noetherian.*

Proof. Our argument will be a modification of the argument used in the proof of Hilbert's theorem for polynomials. We shall consider elements of lowest degree instead of elements of highest degree.

Let $\mathfrak{A}$ be an ideal of $A[[X]]$. We let $\mathfrak{a}_i$ be the set of elements $a \in A$ such that a is the coefficient of X^i in a power series

$$aX^i + \text{terms of higher degree}$$

lying in $\mathfrak{A}$. Then $\mathfrak{a}_i$ is an ideal of A, and $\mathfrak{a}_i \subset \mathfrak{a}_{i+1}$ (the proof of this assertion being the same as for polynomials). The ascending chain of ideals stops:

$$\mathfrak{a}_0 \subset \mathfrak{a}_1 \subset \mathfrak{a}_2 \subset \cdots \subset \mathfrak{a}_r = \mathfrak{a}_{r+1} = \cdots$$

As before, let a_{ij} ($i = 0, \ldots, r$ and $j = 1, \ldots, n_i$) be generators for the ideals $\mathfrak{a}_i$, and let f_{ij} be power series in A having a_{ij} as beginning coefficient. Given $f \in \mathfrak{A}$, starting with a term of degree d, say $d \leqq r$, we can find elements $c_1, \ldots, c_{n_d} \in A$ such that

$$f - c_1 f_{d1} - \cdots - c_{n_d} f_{dn_d}$$

starts with a term of degree $\geqq d + 1$. Proceeding inductively, we may assume that $d > r$. We then use a linear combination

$$f - c_1^{(d)} X^{d-r} f_{r1} - \cdots - c_{n_r}^{(d)} X^{d-r} f_{rn_r}$$

to get a power series starting with a term of degree $\geqq d + 1$. In this way, if we start with a power series of degree $d > r$, then it can be expressed as a linear combination of $f_{r1}, \ldots, f_{rn_r}$ by means of the coefficients

$$g_1(X) = \sum_{\nu=d}^{\infty} c_1^{(\nu)} X^{\nu-r}, \ldots, g_{n_r}(X) = \sum_{\nu=d}^{\infty} c_{n_r}^{(\nu)} X^{\nu-r},$$

and we see that the f_{ij} generate our ideal $\mathfrak{A}$, as was to be shown.

Corollary 9.5. *If A is a Noetherian commutative ring, or a field, then $A[[X_1, \ldots, X_n]]$ is Noetherian.*

Examples. Power series in one variable are at the core of the theory of functions of one complex variable, and similarly for power series in several variables in the higher-dimensional case. See for instance [Gu 90].

Weierstrass polynomials occur in several contexts. First, they can be used to reduce questions about power series to questions about polynomials, in studying analytic sets. See for instance [GrH 78], Chapter 0. In a number-

theoretic context, such polynomials occur as characteristic polynomials in the Iwasawa theory of cyclotomic fields. Cf. [La 90], starting with Chapter 5.

Power series can also be used as generating functions. Suppose that to each positive integer n we associate a number $a(n)$. Then the **generating function** is the power series $\sum a(n)t^n$. In significant cases, it turns out that this function represents a rational function, and it may be a major result to prove that this is so.

For instance in Chapter X, §6 we shall consider a Poincaré series, associated with the length of modules. Similarly, in topology, consider a topological space X such that its homology groups (say) are finite dimensional over a field k of coefficients. Let $h_n = \dim H_n(X, k)$, where H_n is the n-th homology group. The **Poincaré series** is defined to be the generating series

$$P_X(t) = \sum h_n t^n.$$

Examples arise in the theory of dynamical systems. One considers a mapping $T: X \to X$ from a space X into itself, and we let N_n be the number of fixed points of the n-th iterate $T^n = T \circ T \circ \cdots \circ T$ (n times). The generating function is $\sum N_n t^n$. Because of the number of references I give here, I list them systematically at the end of the section. See first Artin–Mazur [ArM 65]; a proof by Manning of a conjecture of Smale [Ma 71]; and Shub's book [Sh 87], especially Chapter 10, Corollary 10.42 (Manning's theorem).

For an example in algebraic geometry, let V be an algebraic variety defined over a finite field k. Let K_n be the extension of k of degree n (in a given algebraic closure). Let N_n be the number of points of V in K_n. One defines the **zeta function** $Z(t)$ as the power series such that $Z(0) = 1$ and

$$Z'/Z(t) = \sum_{n=1}^{\infty} N_n t^{n-1}.$$

Then $Z(t)$ is a rational function (F. K. Schmidt when the dimension of V is 1, and Dwork in higher dimensions). For a discussion and references to the literature, see Appendix C of Hartshorne [Ha 77].

Finally we mention the **partition function** $p(n)$, which is the number of ways a positive integer can be expressed as a sum of positive integers. The generating function was determined by Euler to be

$$1 + \sum_{n=1}^{\infty} p(n)t^n = \prod_{n=1}^{\infty} (1 - t^n)^{-1}.$$

See for instance Hardy and Wright [HardW 71], Chapter XIX. The generating series for the partition function is related to the power series usually expressed in terms of a variable q, namely

$$\Delta = q \prod_{n=1}^{\infty} (1 - q^n)^{24} = \sum_{n=1}^{\infty} \tau(n) q^n,$$

which is the generating series for the **Ramanujan function** $\tau(n)$. The power series for Δ is also the expansion of a function in the theory of modular functions. For an introduction, see Serre's book [Se 73], last chapter, and books on elliptic functions, e.g. mine. We shall mention one application of the power series for Δ in the Galois theory chapter.

Generating power series also occur in K-theory, topological and algebraic geometric, as in Hirzebruch's formalism for the Riemann–Roch theorem and its extension by Grothendieck. See Atiyah [At 67], Hirzebruch [Hi 66], and [FuL 86]. I have extracted some formal elementary aspects having directly to do with power series in Exercises 21–27, which can be viewed as basic examples. See also Exercises 31–34 of the next chapter.

Bibliography

[ArM 65] M. ARTIN and B. MAZUR, On periodic points, *Ann. Math.* (2) **81** (1965) pp. 89–99

[At 67] M. ATIYAH, *K-Theory*, Addison-Wesley 1991 (reprinted from the Benjamin Lecture Notes, 1967)

[FuL 85] W. FULTON and S. LANG, *Riemann–Roch Algebra*, Springer-Verlag, New York, 1985

[GrH 78] P. GRIFFITHS and J. HARRIS, *Principles of Algebraic Geometry*, Wiley–Interscience, New York, 1978

[Gu 90] R. GUNNING, *Introduction to Holomorphic Functions of Several Variables*, Vol. II: *Local Theory*, Wadsworth and Brooks/Cole, 1990

[HardW 71] G. H. HARDY and E. M. WRIGHT, *An Introduction to the Theory of Numbers*, Oxford University Press, Oxford, UK, 1938–1971 (several editions)

[Hart 77] R. HARTSHORNE, *Algebraic Geometry*, Springer-Verlag, New York, 1977

[Hi 66] F. HIRZEBRUCH, *Topological Methods in Algebraic Geometry*, Springer-Verlag, New York, 1966 (translated and expanded from the original German, 1956)

[La 90] S. LANG, *Cyclotomic Fields*, I and II, Springer-Verlag, New York, 1990, combined edition of the original editions, 1978, 1980

[Ma 71] A. MANNING, Axiom A diffeomorphisms have rational zeta functions, *Bull. Lond. Math. Soc.* **3** (1971) pp. 215–220

[Se 73] J. P. SERRE, *A Course in Arithmetic*, Springer-Verlag, New York, 1973

[Sh 87] M. SHUB, *Global Stability of Dynamical Systems*, Springer-Verlag, New York, 1987

EXERCISES

1. Let k be a field and $f(X) \in k[X]$ a non-zero polynomial. Show that the following conditions are equivalent:
   (a) The ideal $(f(X))$ is prime.
   (b) The ideal $(f(X))$ is maximal.
   (c) $f(X)$ is irreducible.

2. (a) State and prove the analogue of Theorem 5.2 for the rational numbers.
   (b) State and prove the analogue of Theorem 5.3 for positive integers.

3. Let f be a polynomial in one variable over a field k. Let X, Y be two variables. Show that in $k[X, Y]$ we have a "Taylor series" expansion

$$f(X + Y) = f(X) + \sum_{i=1}^{n} \varphi_i(X) Y^i,$$

   where $\varphi_i(X)$ is a polynomial in X with coefficients in k. If k has characteristic 0, show that

$$\varphi_i(X) = \frac{D^i f(X)}{i!}.$$

4. Generalize the preceding exercise to polynomials in several variables (introduce partial derivatives and show that a finite Taylor expansion exists for a polynomial in several variables).

5. (a) Show that the polynomials $X^4 + 1$ and $X^6 + X^3 + 1$ are irreducible over the rational numbers.
   (b) Show that a polynomial of degree 3 over a field is either irreducible or has a root in the field. Is $X^3 - 5X^2 + 1$ irreducible over the rational numbers?
   (c) Show that the polynomial in two variables $X^2 + Y^2 - 1$ is irreducible over the rational numbers. Is it irreducible over the complex numbers?

6. Prove the integral root test of §3.

7. (a) Let k be a finite field with $q = p^m$ elements. Let $f(X_1, \ldots, X_n)$ be a polynomial in $k[X]$ of degree d and assume $f(0, \ldots, 0) = 0$. An element $(a_1, \ldots, a_n) \in k^{(n)}$ such that $f(a) = 0$ is called a zero of f. If $n > d$, show that f has at least one other zero in $k^{(n)}$. [*Hint*: Assume the contrary, and compare the degrees of the reduced polynomial belonging to

$$1 - f(X)^{q-1}$$

   and $(1 - X_1^{q-1}) \cdots (1 - X_n^{q-1})$. The theorem is due to Chevalley.]
   (b) Refine the above results by proving that the number N of zeros of f in $k^{(n)}$ is $\equiv 0 \pmod{p}$, arguing as follows. Let i be an integer $\geqq 1$. Show that

$$\sum_{x \in k} x^i = \begin{cases} q - 1 = -1 & \text{if } q - 1 \text{ divides } i, \\ 0 & \text{otherwise.} \end{cases}$$

   Denote the preceding function of i by $\psi(i)$. Show that

$$N \equiv \sum_{x \in k^{(n)}} (1 - f(x)^{q-1})$$

and for each n-tuple $(i_1, \ldots, i_n)$ of integers $\geqq 0$ that

$$\sum_{x \in k^{(n)}} x_1^{i_1} \cdots x_n^{i_n} = \psi(i_1) \cdots \psi(i_n).$$

Show that both terms in the sum for N above yield $0 \bmod p$. (The above argument is due to Warning.)

(c) Extend Chevalley's theorem to r polynomials $f_1, \ldots, f_r$ of degrees $d_1, \ldots, d_r$ respectively, in n variables. If they have no constant term and $n > \sum d_i$, show that they have a non-trivial common zero.

(d) Show that an arbitrary function $f: k^{(n)} \to k$ can be represented by a polynomial. (As before, k is a finite field.)

8. Let A be a commutative entire ring and X a variable over A. Let $a, b \in A$ and assume that a is a unit in A. Show that the map $X \mapsto aX + b$ extends to a unique automorphism of $A[X]$ inducing the identity on A. What is the inverse automorphism?

9. Show that every automorphism of $A[X]$ inducing the identity on A is of the type described in Exercise 8.

10. Let K be a field, and $K(X)$ the quotient field of $K[X]$. Show that every automorphism of $K(X)$ which induces the identity on K is of type

$$X \mapsto \frac{aX + b}{cX + d}$$

with $a, b, c, d \in K$ such that $(aX + b)/(cX + d)$ is not an element of K, or equivalently, $ad - bc \neq 0$.

11. Let A be a commutative entire ring and let K be its quotient field. We show here that some formulas from calculus have a purely algebraic setting. Let $D: A \to A$ be a **derivation**, that is an additive homomorphism satisfying the rule for the derivative of a product, namely

$$D(xy) = xDy + yDx \qquad \text{for} \quad x, y \in A.$$

(a) Prove that D has a unique extension to a derivation of K into itself, and that this extension satisfies the rule

$$D(x/y) = \frac{yDx - xDy}{y^2}$$

for $x, y \in A$ and $y \neq 0$. [Define the extension by this formula, prove that it is independent of the choice of x, y to write the fraction x/y, and show that it is a derivation having the original value on elements of A.]

(b) Let $L(x) = Dx/x$ for $x \in K^*$. Show that $L(xy) = L(x) + L(y)$. The homomorphism L is called the **logarithmic derivative**.

(c) Let D be the standard derivative in the polynomial ring $k[X]$ over a field k. Let $R(X) = c \prod (X - \alpha_i)^{m_i}$ with $\alpha_i \in k$, $c \in k$, and $m_i \in \mathbf{Z}$, so $R(X)$ is a rational

function. Show that

$$R'/R = \sum \frac{m_i}{X - \alpha_i}.$$

12. (a) If $f(X) = aX^2 + bX + c$, show that the discriminant of f is $b^2 - 4ac$.
    (b) If $f(X) = a_0X^3 + a_1X^2 + a_2X + a_3$, show that the discriminant of f is

$$a_1^2a_2^2 - 4a_0a_2^3 - 4a_1^3a_3 - 27a_0^2a_3^2 + 18a_0a_1a_2a_3.$$

    (c) Let $f(X) = (X - t_1) \cdots (X - t_n)$. Show that

$$D_f = (-1)^{n(n-1)/2} \prod_{i=1}^{n} f'(t_i).$$

13. Polynomials will be taken over an algebraically closed field of characteristic 0.
    (a) Prove

Davenport's theorem. *Let $f(t)$, $g(t)$ be polynomials such that $f^3 - g^2 \neq 0$. Then*

$$\deg(f^3 - g^2) \geqq \tfrac{1}{2} \deg f + 1.$$

Or put another way, let $h = f^3 - g^2$ and assume $h \neq 0$. Then

$$\deg f \leqq 2 \deg h - 2.$$

To do this, first assume f, g relatively prime and apply Mason's theorem. In general, proceed as follows.

    (b) *Let A, B, f, g be polynomials such that Af, Bg are relatively prime $\neq 0$. Let $h = Af^3 + Bg^2$. Then*

$$\deg f \leqq \deg A + \deg B + 2 \deg h - 2.$$

This follows directly from Mason's theorem. Then starting with f, g not necessarily relatively prime, start factoring out common factors until no longer possible, to effect the desired reduction. When I did it, I needed to do this step three times, so don't stop until you get it.

    (c) Generalize (b) to the case of $f^m - g^n$ for arbitrary positive integer exponents m and n.

14. Prove that the generalized Szpiro conjecture implies the *abc* conjecture.

15. Prove that the *abc* conjecture implies the following conjecture: There are infinitely many primes p such that $2^{p-1} \not\equiv 1 \bmod p^2$. [Cf. the reference [Sil 88] and [La 90] at the end of §7.]

16. Let w be a complex number, and let $c = \max(1, |w|)$. Let F, G be non-zero polynomials in one variable with complex coefficients, of degrees d and d' respectively, such that $|F|, |G| \geqq 1$. Let R be their resultant. Then

$$|R| \leqq c^{d+d'}[|F(w)| + |G(w)|]|F|^{d'}|G|^{d}(d + d')^{d+d'}.$$

(We denote by $|F|$ the maximum of the absolute values of the coefficients of F.)

17. Let d be an integer $\geqq 3$. Prove the existence of an irreducible polynomial of degree d over $\mathbf{Q}$, having precisely $d - 2$ real roots, and a pair of complex conjugate roots. Use the following construction. Let $b_1, \ldots, b_{d-2}$ be distinct

integers, and let a be an integer > 0. Let

$$g(X) = (X^2 + a)(X - b_1) \cdots (X - b_{d-2}) = X^d + c_{d-1}X^{d-1} + \cdots + c_0.$$

Observe that $c_i \in \mathbf{Z}$ for all i. Let p be a prime number, and let

$$g_n(X) = g(X) + \frac{p}{p^{dn}}$$

so that g_n converges to g (i.e. the coefficients of g_n converge to the coefficients of g).

(a) Prove that g_n has precisely $d - 2$ real roots for n sufficiently large. (You may use a bit of calculus, or use whatever method you want.)

(b) Prove that g_n is irreducible over $\mathbf{Q}$.

Integral-valued polynomials

18. Let $P(X) \in \mathbf{Q}[X]$ be a polynomial in one variable with rational coefficients. It may happen that $P(n) \in \mathbf{Z}$ for all sufficiently large integers n without necessarily P having integer coefficients.

(a) Give an example of this.

(b) Assume that P has the above property. Prove that there are integers $c_0, c_1, \ldots, c_r$ such that

$$P(X) = c_0\binom{X}{r} + c_1\binom{X}{r-1} + \cdots + c_r,$$

where

$$\binom{X}{r} = \frac{1}{r!}X(X-1)\cdots(X-r+1)$$

is the binomial coefficient function. In particular, $P(n) \in \mathbf{Z}$ for all n. Thus we may call P **integral valued**.

(c) Let $f: \mathbf{Z} \to \mathbf{Z}$ be a function. Assume that there exists an integral valued polynomial Q such that the difference function Δf defined by

$$(\Delta f)(n) = f(n) - f(n-1)$$

is equal to $Q(n)$ for all n sufficiently large positive. Show that there exists an integral-valued polynomial P such that $f(n) = P(n)$ for all n sufficiently large.

Exercises on symmetric functions

19. (a) Let $X_1, \ldots, X_n$ be variables. Show that any homogeneous polynomial in $\mathbf{Z}[X_1, \ldots, X_n]$ of degree $> n(n-1)$ lies in the ideal generated by the elementary symmetric functions $s_1, \ldots, s_n$.

(b) With the same notation show that $\mathbf{Z}[X_1, \ldots, X_n]$ is a free $\mathbf{Z}[s_1, \ldots, s_n]$ module with basis the monomials

$$X^{(r)} = X_1^{r_1} \cdots X_n^{r_n}$$

with $0 \leqq r_i \leqq n - i$.

(c) Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ be two independent sets of variables. Let $s_1, \ldots, s_n$ be the elementary symmetric functions of X and $s'_1, \ldots, s'_m$ the elementary symmetric functions of Y (using vector vector notation). Show that $\mathbf{Z}[X, Y]$ is free over $\mathbf{Z}[s, s']$ with basis $X^{(r)}Y^{(q)}$, and the exponents (r), (q) satisfying inequalities as in (b).

(d) Let I be an ideal in $\mathbf{Z}[s, s']$. Let J be the ideal generated by I in $\mathbf{Z}[X, Y]$. Show that

$$J \cap \mathbf{Z}[s, s'] = I.$$

20. Let A be a commutative ring. Let t be a variable. Let

$$f(t) = \sum_{i=0}^{m} a_i t^i \qquad \text{and} \qquad g(t) = \sum_{i=0}^{n} b_i t^i$$

be polynomials whose constant terms are $a_0 = b_0 = 1$. If

$$f(t)g(t) = 1,$$

show that there exists an integer N $\big(= (m + n)(m + n - 1)\big)$ such that any monomial

$$a_1^{r_1} \cdots a_n^{r_n}$$

with $\sum jr_j > N$ is equal to 0. [*Hint*: Replace the a's and b's by variables. Use Exercise 19(b) to show that any monomial $M(a)$ of weight $> N$ lies in the ideal I generated by the elements

$$c_k = \sum_{i=0}^{k} a_i b_{k-i}$$

(letting $a_0 = b_0 = 1$). Note that c_k is the k-th elementary symmetric function of the $m + n$ variables (X, Y).]

[*Note*: For some interesting contexts involving symmetric functions, see Cartier's talk at the Bourbaki Seminar, 1982–1983.]

λ-rings

The following exercises start a train of thought which will be pursued in Exercise 33 of Chapter V; Exercises 22–24 of Chapter XVIII; and Chapter XX, §3. These originated to a large extent in Hirzebruch's Riemann–Roch theorem and its extension by Grothendieck who defined λ-rings in general.

Let K be a commutative ring. By **λ-operations** we mean a family of mappings

$$\lambda^i : K \to K$$

for each integer $i \geqq 0$ satisfying the relations for all $x \in K$:

$$\lambda^0(x) = 1, \qquad \lambda^1(x) = x,$$

and for all integers $n \geqq 0$, and $x, y \in K$,

$$\lambda^n(x + y) = \sum_{i=0}^{n} \lambda^i(x)\lambda^{n-i}(y).$$

The reader will meet examples of such operations in the chapter on the alternating and symmetric products, but the formalism of such operations depends only on the above relations, and so can be developed here in the context of formal power series. Given a λ-operation, in which case we also say that K is a **λ-ring**, we define the power series

$$\lambda_t(x) = \sum_{i=0}^{\infty} \lambda^i(x)t^i.$$

Prove the following statements.

21. The map $x \mapsto \lambda_t(x)$ is a homomorphism from the additive group of K into the multiplicative group of power series $1 + tK[[t]]$ whose constant term is equal to 1. Conversely, any such homomorphism such that $\lambda_t(x) = 1 + xt +$ higher terms gives rise to λ-operations.

22. Let $s = at +$ higher terms be a power series in $K[[t]]$ such that a is a unit in K. Show that there is a power series

$$t = g(s) = \sum b_i s^i \qquad \text{with} \quad b_i \in K.$$

Show that any power series $f(t) \in K[[t]]$ can be written in the form $h(s)$ for some other power series with coefficients in K.

Given a λ-operation on K, define the corresponding **Grothendieck power series**

$$\gamma_t(x) = \lambda_{t/(1-t)}(x) = \lambda_s(x)$$

where $s = t/(1-t)$. Then the map

$$x \mapsto \gamma_t(x)$$

is a homomorphism as before. We define $\gamma^i(x)$ by the relation

$$\gamma_t(x) = \sum \gamma^i(x)t^i.$$

Show that γ satisfies the following properties.

23. (a) For every integer $n \geqq 0$ we have

$$\gamma^n(x + y) = \sum_{i=0}^{n} \gamma^i(x)\gamma^{n-i}(y).$$

(b) $\gamma_t(1) = 1/(1-t)$.

(c) $\gamma_t(-1) = 1 - t$.

24. Assume that $\lambda^i u = 0$ for $i > 1$. Show:

(a) $\gamma_t(u - 1) = 1 + (u-1)t$.

(b) $\gamma_t(1 - u) = \sum_{i=0}^{\infty} (1-u)^i t^i$.

25. **Bernoulli numbers.** Define the Bernoulli numbers B_k as the coefficients in the power series

$$F(t) = \frac{t}{e^t - 1} = \sum_{k=0}^{\infty} B_k \frac{t^k}{k!}.$$

Of course, $e^t = \sum t^n/n!$ is the standard power series with rational coefficients $1/n!$. Prove:

(a) $B_0 = 1$, $B_1 = -\frac{1}{2}$, $B_2 = \frac{1}{6}$.

(b) $F(-t) = t + F(t)$, and $B_k = 0$ if k is odd $\neq 1$.

26. **Bernoulli polynomials.** Define the Bernoulli polynomials $\mathbf{B}_k(X)$ by the power series expansion

$$F(t, X) = \frac{te^{tX}}{e^t - 1} = \sum_{k=0}^{\infty} \mathbf{B}_k(X)\frac{t^k}{k!}.$$

It is clear that $B_k = \mathbf{B}_k(0)$, so the Bernoulli numbers are the constant terms of the Bernoulli polynomials. Prove:

(a) $\mathbf{B}_0(X) = 1$, $\mathbf{B}_1(X) = X - \frac{1}{2}$, $\mathbf{B}_2(X) = X^2 - X + \frac{1}{6}$.

(b) For each positive integer N,

$$\mathbf{B}_k(X) = N^{k-1} \sum_{a=0}^{N-1} \mathbf{B}_k\left(\frac{X + a}{N}\right).$$

(c) $\mathbf{B}_k(X) = X^k - \frac{1}{2}kX^{k-1} +$ lower terms.

(d) $F(t, X + 1) - F(t, X) = te^{Xt} = t\sum X^k \dfrac{t^k}{k!}$.

(e) $\mathbf{B}_k(X + 1) - \mathbf{B}_k(X) = kX^{k-1}$ for $k \geqq 1$.

27. Let N be a positive integer and let f be a function on $\mathbf{Z}/N\mathbf{Z}$. Form the power series

$$F_f(t, X) = \sum_{a=0}^{N-1} f(a)\frac{te^{(a+X)t}}{e^{Nt} - 1}.$$

Following Leopoldt, define the **generalized Bernoulli polynomials** relative to the function f by

$$F_f(t, X) = \sum_{k=0}^{\infty} \mathbf{B}_{k,f}(X)\frac{t^k}{k!}.$$

In particular, the constant term of $\mathbf{B}_{k,f}(X)$ is defined to be the **generalized Bernoulli number** $B_{k,f} = \mathbf{B}_{k,f}(0)$ introduced by Leopoldt in cyclotomic fields. Prove:

(a) $F_f(t, X + k) = e^{kt}F_f(t, X)$.

(b) $F_f(t, X + N) - F_f(t, X) = (e^{Nt} - 1)F_f(t, X)$.

(c) $\dfrac{1}{k}[\mathbf{B}_{k,f}(X + N) - \mathbf{B}_{k,f}(X)] = \sum_{a=0}^{N-1} f(a)(a + X)^{k-1}$.

(d) $\mathbf{B}_{k,f}(X) = \sum_{i=0}^{k} \binom{k}{i} B_{i,f}X^{n-i}$

$$= B_{k,f} + kB_{k-1,f}X + \cdots + kB_{1,f}X^{k-1} + B_{0,f}X^k.$$

Note. The exercises on Bernoulli numbers and polynomials are designed not only to give examples for the material in the text, but to show how this material leads into major areas of mathematics: in topology and algebraic geometry centering

around Riemann–Roch theorems; analytic and algebraic number theory, as in the theory of the zeta functions and the theory of modular forms, cf. my *Introduction to Modular Forms*, Springer-Verlag, New York, 1976, Chapters XIV and XV; my *Cyclotomic Fields*, I and II, Springer-Verlag, New York, 1990, Chapter 2, §2; Kubert–Lang's *Modular Units*, Springer-Verlag, New York, 1981; etc.

Further Comments, 1996–2001. I was informed by Umberto Zannier that what has been called Mason's theorem was proved three years earlier by Stothers [Sto 81], Theorem 1.1. Zannier himself has published some results on Davenport's theorem [Za 95], without knowing of the paper by Stothers, using a method similar to that of Stothers, and rediscovering some of Stothers' results, but also going beyond. Indeed, Stothers uses the "Belyi method" belonging to algebraic geometry, and increasingly appearing as a fundamental tool. Mason gave a very elementary proof, accessible at the basic level of algebra. An even shorter and very elegant proof of the Mason-Stothers theorem was given by Noah Snyder [Sny 00]. I am much indebted to Snyder for showing me that proof before publication, and I reproduced it in [La 99b]. But I recommend looking at Snyder's version.

[La 99b] S. LANG, *Math Talks for Undergraduates*, Springer Verlag 1999

[Sny 00] N. SNYDER, An alternate proof of Mason's theorem, *Elemente der Math.* **55** (2000) pp. 93–94

[Sto 81] W. STOTHERS, Polynomial identities and hauptmoduln, *Quart. J. Math. Oxford* (2) **32** (1981) pp. 349–370

[Za 95] U. ZANNIER, On Davenport's bound for the degree of $f^3 - g^2$ and Riemann's existence theorem, *Acta Arithm.* **LXXI.2** (1995) pp. 107–137

Part Two

ALGEBRAIC EQUATIONS

This part is concerned with the solutions of algebraic equations, in one or several variables. This is the recurrent theme in every chapter of this part, and we lay the foundations for all further studies concerning such equations.

Given a subring A of a ring B, and a finite number of polynomials $f_1, \ldots, f_n$ in $A[X_1, \ldots, X_n]$, we are concerned with the n-tuples

$$(b_1, \ldots, b_n) \in B^{(n)}$$

such that

$$f_i(b_1, \ldots, b_n) = 0$$

for $i = 1, \ldots, r$. For suitable choices of A and B, this includes the general problem of diophantine analysis when A, B have an "arithmetic" structure.

We shall study various cases. We begin by studying roots of one polynomial in one variable over a field. We prove the existence of an algebraic closure, and emphasize the role of irreducibility.

Next we study the group of automorphisms of algebraic extensions of a field, both intrinsically and as a group of permutations of the roots of a polynomial. We shall mention some major unsolved problems along the way.

It is also necessary to discuss extensions of a ring, to give the possibility of analyzing families of extensions. The ground work is laid in Chapter VII.

In Chapter IX, we come to the zeros of polynomials in several variables, essentially over algebraically closed fields. But again, it is advantageous to

consider polynomials over rings, especially $\mathbf{Z}$, since in projective space, the conditions that homogeneous polynomials have a non-trivial common zero can be given universally over $\mathbf{Z}$ in terms of their coefficients.

Finally we impose additional structures like those of reality, or metric structures given by absolute values. Each one of these structures gives rise to certain theorems describing the structure of the solutions of equations as above, and especially proving the existence of solutions in important cases.

CHAPTER V

Algebraic Extensions

In this first chapter concerning polynomial equations, we show that given a polynomial over a field, there always exists some extension of the field where the polynomial has a root, and we prove the existence of an algebraic closure. We make a preliminary study of such extensions, including the automorphisms, and we give algebraic extensions of finite fields as examples.

§1. FINITE AND ALGEBRAIC EXTENSIONS

Let F be a field. If F is a subfield of a field E, then we also say that E is an **extension field** of F. We may view E as a vector space over F, and we say that E is a **finite** or **infinite** extension of F according as the dimension of this vector space is finite or infinite.

Let F be a subfield of a field E. An element α of E is said to be **algebraic** over F if there exist elements $a_0, \ldots, a_n$ $(n \geqq 1)$ of F, not all equal to 0, such that

$$a_0 + a_1\alpha + \cdots + a_n\alpha^n = 0.$$

If $\alpha \neq 0$, and α is algebraic, then we can always find elements a_i as above such that $a_0 \neq 0$ (factoring out a suitable power of α).

Let X be a variable over F. We can also say that α is algebraic over F if the homomorphism

$$F[X] \to E$$

S. Lang, *Algebra*
© Springer Science+Business Media LLC 2002

which is the identity on F and maps X on α has a non-zero kernel. In that case the kernel is an ideal which is principal, generated by a single polynomial $p(X)$, which we may assume has leading coefficient 1. We then have an isomorphism

$$F[X]/(p(X)) \approx F[\alpha],$$

and since $F[\alpha]$ is entire, it follows that $p(X)$ is irreducible. Having normalized $p(X)$ so that its leading coefficient is 1, we see that $p(X)$ is uniquely determined by α and will be called THE irreducible polynomial of α over F. We sometimes denote it by $\mathrm{Irr}(\alpha, F, X)$.

An extension E of F is said to be **algebraic** if every element of E is algebraic over F.

Proposition 1.1. *Let E be a finite extension of F. Then E is algebraic over F.*

Proof. Let $\alpha \in E$, $\alpha \neq 0$. The powers of α,

$$1, \alpha, \alpha^2, \ldots, \alpha^n,$$

cannot be linearly independent over F for all positive integers n, otherwise the dimension of E over F would be infinite. A linear relation between these powers shows that α is algebraic over F.

Note that the converse of Proposition 1.1 is not true; there exist infinite algebraic extensions. We shall see later that the subfield of the complex numbers consisting of all algebraic numbers over $\mathbf{Q}$ is an infinite extension of $\mathbf{Q}$.

If E is an extension of F, we denote by

$$[E : F]$$

the dimension of E as vector space over F. It may be infinite.

Proposition 1.2. *Let k be a field and $F \subset E$ extension fields of k. Then*

$$[E : k] = [E : F][F : k].$$

If $\{x_i\}_{i \in I}$ is a basis for F over k and $\{y_j\}_{j \in J}$ is a basis for E over F, then $\{x_i y_j\}_{(i,j) \in I \times J}$ is a basis for E over k.

Proof. Let $z \in E$. By hypothesis there exist elements $\alpha_j \in F$, almost all $\alpha_j = 0$, such that

$$z = \sum_{j \in J} \alpha_j y_j.$$

For each $j \in J$ there exist elements $b_{ji} \in k$, almost all of which are equal to 0, such that

$$\alpha_j = \sum_{i \in I} b_{ji} x_i,$$

and hence

$$z = \sum_j \sum_i b_{ji} x_i y_j.$$

This shows that $\{x_i y_j\}$ is a family of generators for E over k. We must show that it is linearly independent. Let $\{c_{ij}\}$ be a family of elements of k, almost all of which are 0, such that

$$\sum_j \sum_i c_{ij} x_i y_j = 0.$$

Then for each j,

$$\sum_i c_{ij} x_i = 0$$

because the elements y_j are linearly independent over F. Finally $c_{ij} = 0$ for each i because $\{x_i\}$ is a basis of F over k, thereby proving our proposition.

Corollary 1.3. *The extension E of k is finite if and only if E is finite over F and F is finite over k.*

As with groups, we define a **tower** of fields to be a sequence

$$F_1 \subset F_2 \subset \cdots \subset F_n$$

of extension fields. The tower is called **finite** if and only if each step is finite.

Let k be a field, E an extension field, and $\alpha \in E$. We denote by $k(\alpha)$ the smallest subfield of E containing both k and α. It consists of all quotients $f(\alpha)/g(\alpha)$, where f, g are polynomials with coefficients in k and $g(\alpha) \neq 0$.

Proposition 1.4. *Let α be algebraic over k. Then $k(\alpha) = k[\alpha]$, and $k(\alpha)$ is finite over k. The degree $[k(\alpha) : k]$ is equal to the degree of* $\operatorname{Irr}(\alpha, k, X)$.

Proof. Let $p(X) = \operatorname{Irr}(\alpha, k, X)$. Let $f(X) \in k[X]$ be such that $f(\alpha) \neq 0$. Then $p(X)$ does not divide $f(X)$, and hence there exist polynomials $g(X)$, $h(X) \in k[X]$ such that

$$g(X)p(X) + h(X)f(X) = 1.$$

From this we get $h(\alpha)f(\alpha) = 1$, and we see that $f(\alpha)$ is invertible in $k[\alpha]$. Hence $k[\alpha]$ is not only a ring but a field, and must therefore be equal to $k(\alpha)$. Let $d = \deg p(X)$. The powers

$$1, \alpha, \ldots, \alpha^{d-1}$$

are linearly independent over k, for otherwise suppose

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

with $a_i \in k$, not all $a_i = 0$. Let $g(X) = a_0 + \cdots + a_{d-1}X^{d-1}$. Then $g \neq 0$ and $g(\alpha) = 0$. Hence $p(X)$ divides $g(X)$, contradiction. Finally, let $f(\alpha) \in k[\alpha]$, where $f(X) \in k[X]$. There exist polynomials $q(X)$, $r(X) \in k[X]$ such that $\deg r < d$ and

$$f(X) = q(X)p(X) + r(X).$$

Then $f(\alpha) = r(\alpha)$, and we see that $1, \alpha, \ldots, \alpha^{d-1}$ generate $k[\alpha]$ as a vector space over k. This proves our proposition.

Let E, F be extensions of a field k. If E and F are contained in some field L then we denote by EF the smallest subfield of L containing both E and F, and call it the **compositum** of E and F, in L. If E, F are not given as embedded in a common field L, then we cannot define the compositum.

Let k be a subfield of E and let $\alpha_1, \ldots, \alpha_n$ be elements of E. We denote by

$$k(\alpha_1, \ldots, \alpha_n)$$

the smallest subfield of E containing k and $\alpha_1, \ldots, \alpha_n$. Its elements consist of all quotients

$$\frac{f(\alpha_1, \ldots, \alpha_n)}{g(\alpha_1, \ldots, \alpha_n)}$$

where f, g are polynomials in n variables with coefficients in k, and

$$g(\alpha_1, \ldots, \alpha_n) \neq 0.$$

Indeed, the set of such quotients forms a field containing k and $\alpha_1, \ldots, \alpha_n$. Conversely, any field containing k and

$$\alpha_1, \ldots, \alpha_n$$

must contain these quotients.

We observe that E is the union of all its subfields $k(\alpha_1, \ldots, \alpha_n)$ as $(\alpha_1, \ldots, \alpha_n)$ ranges over finite subfamilies of elements of E. We could define the *compositum of an arbitrary subfamily of subfields of a field* L as the smallest subfield containing all fields in the family. We say that E is **finitely generated** over k if there is a finite family of elements $\alpha_1, \ldots, \alpha_n$ of E such that

$$E = k(\alpha_1, \ldots, \alpha_n).$$

We see that E is the compositum of all its finitely generated subfields over k.

Proposition 1.5. *Let E be a finite extension of k. Then E is finitely generated.*

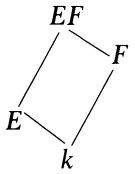
Proof. Let $\{\alpha_1, \ldots, \alpha_n\}$ be a basis of E as vector space over k. Then certainly

$$E = k(\alpha_1, \ldots, \alpha_n).$$

If $E = k(\alpha_1, \ldots, \alpha_n)$ is finitely generated, and F is an extension of k, both F, E contained in L, then

$$EF = F(\alpha_1, \ldots, \alpha_n),$$

and EF is finitely generated over F. We often draw the following picture:



Lines slanting up indicate an inclusion relation between fields. We also call the extension EF of F the **translation** of E to F, or also the **lifting** of E to F.

Let α be algebraic over the field k. Let F be an extension of k, and assume $k(\alpha)$, F both contained in some field L. Then α is algebraic over F. Indeed, the irreducible polynomial for α over k has *a fortiori* coefficients in F, and gives a linear relation for the powers of α over F.

Suppose that we have a tower of fields:

$$k \subset k(\alpha_1) \subset k(\alpha_1, \alpha_2) \subset \cdots \subset k(\alpha_1, \ldots, \alpha_n),$$

each one generated from the preceding field by a single element. Assume that each α_i is algebraic over k, $i = 1, \ldots, n$. As a special case of our preceding remark, we note that α_{i+1} is algebraic over $k(\alpha_1, \ldots, \alpha_i)$. Hence each step of the tower is algebraic.

Proposition 1.6. *Let $E = k(\alpha_1, \ldots, \alpha_n)$ be a finitely generated extension of a field k, and assume α_i algebraic over k for each $i = 1, \ldots, n$. Then E is finite algebraic over k.*

Proof. From the above remarks, we know that E can be obtained as the end of a tower each of whose steps is generated by one algebraic element, and is therefore finite by Proposition 1.4. We conclude that E is finite over k by Corollary 1.3, and that it is algebraic by Proposition 1.1.

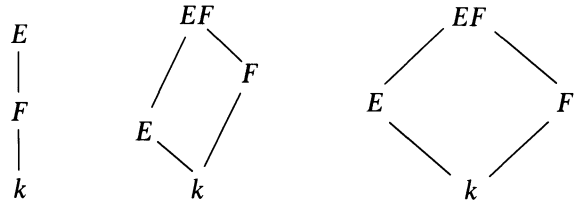
Let $\mathfrak{C}$ be a certain class of extension fields $F \subset E$. We shall say that $\mathfrak{C}$ is **distinguished** if it satisfies the following conditions:

(1) Let $k \subset F \subset E$ be a tower of fields. The extension $k \subset E$ is in $\mathfrak{C}$ if and only if $k \subset F$ is in $\mathfrak{C}$ and $F \subset E$ is in $\mathfrak{C}$.

(2) If $k \subset E$ is in $\mathfrak{C}$, if F is any extension of k, and E, F are both contained in some field, then $F \subset EF$ is in $\mathfrak{C}$.

(3) If $k \subset F$ and $k \subset E$ are in $\mathfrak{C}$ and F, E are subfields of a common field, then $k \subset FE$ is in $\mathfrak{C}$.

The diagrams illustrating our properties are as follows:



(1) (2) (3)

These lattice diagrams of fields are extremely suggestive in handling extension fields.

We observe that (3) follows formally from the first two conditions. Indeed, one views EF over k as a tower with steps $k \subset F \subset EF$.

As a matter of notation, it is convenient to write E/F instead of $F \subset E$ to denote an extension. There can be no confusion with factor groups since we shall never use the notation E/F to denote such a factor group when E is an extension field of F.

Proposition 1.7. *The class of algebraic extensions is distinguished, and so is the class of finite extensions.*

Proof. Consider first the class of finite extensions. We have already proved condition (1). As for (2), assume that E/k is finite, and let F be any extension of k. By Proposition 1.5 there exist elements $\alpha_1, \ldots, \alpha_n \in E$ such that $E = k(\alpha_1, \ldots, \alpha_n)$. Then $EF = F(\alpha_1, \ldots, \alpha_n)$, and hence EF/F is finitely generated by algebraic elements. Using Proposition 1.6 we conclude that EF/F is finite.

Consider next the class of algebraic extensions, and let

$$k \subset F \subset E$$

be a tower. Assume that E is algebraic over k. Then *a fortiori*, F is algebraic over k and E is algebraic over F. Conversely, assume each step in the tower to be algebraic. Let $\alpha \in E$. Then α satisfies an equation

$$a_n\alpha^n + \cdots + a_0 = 0$$

with $a_i \in F$, not all $a_i = 0$. Let $F_0 = k(a_n, \ldots, a_0)$. Then F_0 is finite over k by Proposition 1.6, and α is algebraic over F_0. From the tower

$$k \subset F_0 = k(a_n, \ldots, a_0) \subset F_0(\alpha)$$

and the fact that each step in this tower is finite, we conclude that $F_0(\alpha)$ is finite over k, whence α is algebraic over k, thereby proving that E is algebraic over k and proving condition (1) for algebraic extensions. Condition (2) has already been observed to hold, i.e. an element remains algebraic under lifting, and hence so does an extension.

Remark. It is true that finitely generated extensions form a distinguished class, but one argument needed to prove part of (1) can be carried out only with more machinery than we have at present. Cf. the chapter on transcendental extensions.

§2. ALGEBRAIC CLOSURE

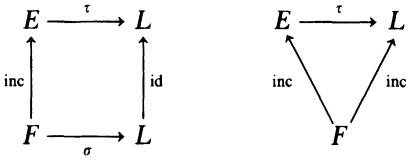
In this and the next section we shall deal with embeddings of a field into another. We therefore define some terminology.

Let E be an extension of a field F and let

$$\sigma: F \to L$$

be an embedding (i.e. an injective homomorphism) of F into L. Then σ induces an isomorphism of F with its image σF, which is sometimes written F^σ. An embedding τ of E in L will be said to be **over** σ if the restriction of τ to F is equal to σ. We also say that τ **extends** σ. If σ is the identity then we say that τ is an embedding of E **over** F.

These definitions could be made in more general categories, since they depend only on diagrams to make sense:



We shall use exponential notation (to avoid parentheses), so we write F^σ instead of σF, and f^σ instead of σf for a polynomial f, applying σ to the coefficients. Cf. Chapter II, §5.

Remark. Let $f(X) \in F[X]$ be a polynomial, and let α be a root of f in E. Say $f(X) = a_0 + \cdots + a_n X^n$. Then

$$0 = f(\alpha) = a_0 + a_1\alpha + \cdots + a_n\alpha^n.$$

If τ extends σ as above, then we see that $\tau\alpha$ is a root of f^σ because

$$0 = \tau(f(\alpha)) = a_0^\sigma + a_1^\sigma(\tau\alpha) + \cdots + a_n^\sigma(\tau\alpha)^n.$$

In our study of embeddings it will also be useful to have a lemma concerning embeddings of algebraic extensions into themselves. For this we note that if $\sigma: E \to L$ is an embedding over k (i.e. inducing the identity on k), then σ can be viewed as a k-homomorphism of vector spaces, because both E, L can be viewed as vector spaces over k. Furthermore σ is injective.

Lemma 2.1. *Let E be an algebraic extension of k, and let $\sigma: E \to E$ be an embedding of E into itself over k. Then σ is an automorphism.*

Proof. Since σ is injective, it will suffice to prove that σ is surjective. Let α be an element of E, let $p(X)$ be its irreducible polynomial over k, and let E' be the subfield of E generated by all the roots of $p(X)$ which lie in E. Then E' is finitely generated, hence is a finite extension of k. Furthermore, σ must map a root of $p(X)$ on a root of $p(X)$, and hence σ maps E' into itself. We can view σ as a k-homomorphism of vector spaces because σ induces the identity on k. Since σ is injective, its image $\sigma(E')$ is a subspace of E' having the same dimension $[E' : k]$. Hence $\sigma(E') = E'$. Since $\alpha \in E'$, it follows that α is in the image of σ, and our lemma is proved.

Let E, F be extensions of a field k, contained in some bigger field L. We can form the ring $E[F]$ generated by the elements of F over E. Then $E[F] = F[E]$, and EF is the quotient field of this ring. It is clear that the elements of $E[F]$ can be written in the form

$$a_1 b_1 + \cdots + a_n b_n$$

with $a_i \in E$ and $b_i \in F$. Hence EF is the field of quotients of these elements.

Lemma 2.2. *Let E_1, E_2 be extensions of a field k, contained in some bigger field E, and let σ be an embedding of E in some field L. Then*

$$\sigma(E_1 E_2) = \sigma(E_1)\sigma(E_2).$$

Proof. We apply σ to a quotient of elements of the above type, say

$$\sigma\left(\frac{a_1 b_1 + \cdots + a_n b_n}{a_1' b_1' + \cdots + a_m' b_m'}\right) = \frac{a_1^\sigma b_1^\sigma + \cdots + a_n^\sigma b_n^\sigma}{a_1'^\sigma b_1'^\sigma + \cdots + a_m'^\sigma b_m'^\sigma},$$

and see that the image is an element of $\sigma(E_1)\sigma(E_2)$. It is clear that the image $\sigma(E_1 E_2)$ is $\sigma(E_1)\sigma(E_2)$.

Let k be a field, $f(X)$ a polynomial of degree $\geqq 1$ in $k[X]$. We consider the problem of finding an extension E of k in which f has a root. If $p(X)$ is an irreducible polynomial in $k[X]$ which divides $f(X)$, then any root of $p(X)$ will also be a root of $f(X)$, so we may restrict ourselves to irreducible polynomials.

Let $p(X)$ be irreducible, and consider the canonical homomorphism

$$\sigma: k[X] \to k[X]/(p(X)).$$

Then σ induces a homomorphism on k, whose kernel is 0, because every nonzero element of k is invertible in k, generates the unit ideal, and 1 does not lie in the kernel. Let ξ be the image of X under σ, i.e. $\xi = \sigma(X)$ is the residue class of X mod $p(X)$. Then

$$p^\sigma(\xi) = p^\sigma(X^\sigma) = \big(p(X)\big)^\sigma = 0.$$

Hence ξ is a root of p^σ, and as such is algebraic over σk. We have now found an extension of σk, namely $\sigma k(\xi)$ in which p^σ has a root.

With a minor set-theoretic argument, we shall have:

Proposition 2.3. *Let k be a field and f a polynomial in $k[X]$ of degree $\geqq 1$. Then there exists an extension E of k in which f has a root.*

Proof. We may assume that $f = p$ is irreducible. We have shown that there exists a field F and an embedding

$$\sigma: k \to F$$

such that p^σ has a root ξ in F. Let S be a set whose cardinality is the same as that of $F - \sigma k$ (= the complement of σk in F) and which is disjoint from k. Let $E = k \cup S$. We can extend $\sigma: k \to F$ to a bijection of E on F. We now define a field structure on E. If $x, y \in E$ we define

$$xy = \sigma^{-1}(\sigma(x)\sigma(y)),$$

$$x + y = \sigma^{-1}(\sigma(x) + \sigma(y)).$$

Restricted to k, our addition and multiplication coincide with the given addition and multiplication of our original field k, and it is clear that k is a subfield of E. We let $\alpha = \sigma^{-1}(\xi)$. Then it is also clear that $p(\alpha) = 0$, as desired.

Corollary 2.4. *Let k be a field and let $f_1, \ldots, f_n$ be polynomials in $k[X]$ of degrees $\geqq 1$. Then there exists an extension E of k in which each f_i has a root, $i = 1, \ldots, n$.*

Proof. Let E_1 be an extension in which f_1 has a root. We may view f_2 as a polynomial over E_1. Let E_2 be an extension of E_1 in which f_2 has a root. Proceeding inductively, our corollary follows at once.

We define a field L to be **algebraically closed** if every polynomial in $L[X]$ of degree $\geqq 1$ has a root in L.

Theorem 2.5. *Let k be a field. Then there exists an algebraically closed field containing k as a subfield.*

Proof. We first construct an extension E_1 of k in which every polynomial in $k[X]$ of degree $\geqq 1$ has a root. One can proceed as follows (Artin). To each polynomial f in $k[X]$ of degree $\geqq 1$ we associate a letter X_f and we let S be the set of all such letters X_f (so that S is in bijection with the set of polynomials in $k[X]$ of degree $\geqq 1$). We form the polynomial ring $k[S]$, and contend that the ideal generated by all the polynomials $f(X_f)$ in $k[S]$ is not the unit ideal. If it is, then there is a finite combination of elements in our ideal which is equal to 1:

$$g_1 f_1(X_{f_1}) + \cdots + g_n f_n(X_{f_n}) = 1$$

with $g_i \in k[S]$. For simplicity, write X_i instead of X_{f_i}. The polynomials g_i will involve actually only a finite number of variables, say $X_1, \ldots, X_N$ (with $N \geqq n$). Our relation then reads

$$\sum_{i=1}^{n} g_i(X_1, \ldots, X_N) f_i(X_i) = 1.$$

Let F be a finite extension in which each polynomial $f_1, \ldots, f_n$ has a root, say α_i is a root of f_i in F, for $i = 1, \ldots, n$. Let $\alpha_i = 0$ for $i > n$. Substitute α_i for X_i in our relation. We get $0 = 1$, contradiction.

Let $\mathfrak{m}$ be a maximal ideal containing the ideal generated by all polynomials $f(X_f)$ in $k[S]$. Then $k[S]/\mathfrak{m}$ is a field, and we have a canonical map

$$\sigma: k[S] \to k[S]/\mathfrak{m}.$$

For any polynomial $f \in k[X]$ of degree $\geqq 1$, the polynomial f^σ has a root in $k[S]/\mathfrak{m}$, which is an extension of σk. Using the same type of set-theoretic argument as in Proposition 2.3, we conclude that there exists an extension E_1 of k in which every polynomial $f \in k[X]$ of degree $\geqq 1$ has a root in E_1.

Inductively, we can form a sequence of fields

$$E_1 \subset E_2 \subset E_3 \subset \cdots \subset E_n \cdots$$

such that every polynomial in $E_n[X]$ of degree $\geqq 1$ has a root in E_{n+1}. Let E be the union of all fields E_n, $n = 1, 2, \ldots$. Then E is naturally a field, for if $x, y \in E$ then there exists some n such that $x, y \in E_n$, and we can take the product or sum xy or $x + y$ in E_n. This is obviously independent of the choice of n such that $x, y \in E_n$, and defines a field structure on E. Every polynomial in $E[X]$ has its coefficients in some subfield E_n, hence a root in E_{n+1}, hence a root in E, as desired.

Corollary 2.6. *Let k be a field. There exists an extension k^a which is algebraic over k and algebraically closed.*

Proof. Let E be an extension of k which is algebraically closed and let k^a be the union of all subextensions of E, which are algebraic over k. Then k^a is algebraic over k. If $\alpha \in E$ and α is algebraic over k^a then α is algebraic over k by Proposition 1.7. If f is a polynomial of degree $\geqq 1$ in $k^a[X]$, then f has a root α in E, and α is algebraic over k^a. Hence α is in k^a and k^a is algebraically closed.

We observe that if L is an algebraically closed field, and $f \in L[X]$ has degree $\geqq 1$, then there exists $c \in L$ and $\alpha_1, \ldots, \alpha_n \in L$ such that

$$f(X) = c(X - \alpha_1) \cdots (X - \alpha_n).$$

Indeed, f has a root α_1 in L, so there exists $g(X) \in L[X]$ such that

$$f(X) = (X - \alpha_1) g(X).$$

If $\deg g \geqq 1$, we can repeat this argument inductively, and express f as a

product of terms $(X - \alpha_i)$ $(i = 1, \ldots, n)$ and an element $c \in L$. Note that c is the leading coefficient of f, i.e.

$$f(X) = cX^n + \text{terms of lower degree.}$$

Hence if the coefficients of f lie in a subfield k of L, then $c \in k$.

Let k be a field and $\sigma: k \to L$ an embedding of k into an algebraically closed field L. We are interested in analyzing the extensions of σ to algebraic extensions E of k. We begin by considering the special case when E is generated by one element.

Let $E = k(\alpha)$ where α is algebraic over k. Let

$$p(X) = \operatorname{Irr}(\alpha, k, X).$$

Let β be a root of p^σ in L. Given an element of $k(\alpha) = k[\alpha]$, we can write it in the form $f(\alpha)$ with some polynomial $f(X) \in k[X]$. We define an extension of σ by mapping

$$f(\alpha) \mapsto f^\sigma(\beta).$$

This is in fact well defined, i.e. independent of the choice of polynomial $f(X)$ used to express our element in $k[\alpha]$. Indeed, if $g(X)$ is in $k[X]$ and such that $g(\alpha) = f(\alpha)$, then $(g - f)(\alpha) = 0$, whence $p(X)$ divides $g(X) - f(X)$. Hence $p^\sigma(X)$ divides $g^\sigma(X) - f^\sigma(X)$, and thus $g^\sigma(\beta) = f^\sigma(\beta)$. It is now clear that our map is a homomorphism inducing σ on k, and that it is an extension of σ to $k(\alpha)$. Hence we get:

Proposition 2.7. *The number of possible extensions of σ to $k(\alpha)$ is $\leqq \deg p$, and is equal to the number of distinct roots of p in k^a.*

This is an important fact, which we shall analyze more closely later. For the moment, we are interested in extensions of σ to arbitrary algebraic extensions of k. We get them by using Zorn's lemma.

Theorem 2.8. *Let k be a field, E an algebraic extension of k, and $\sigma: k \to L$ an embedding of k into an algebraically closed field L. Then there exists an extension of σ to an embedding of E in L. If E is algebraically closed and L is algebraic over σk, then any such extension of σ is an isomorphism of E onto L.*

Proof. Let S be the set of all pairs (F, τ) where F is a subfield of E containing k, and τ is an extension of σ to an embedding of F in L. If (F, τ) and (F', τ') are such pairs, we write $(F, \tau) \leqq (F', \tau')$ if $F \subset F'$ and $\tau' | F = \tau$. Note that S is not empty [it contains (k, σ)], and is inductively ordered: If $\{(F_i, \tau_i)\}$ is a totally ordered subset, we let $F = \bigcup F_i$ and define τ on F to be equal to τ_i on each F_i. Then (F, τ) is an upper bound for the totally ordered subset. Using Zorn's lemma, let (K, λ) be a maximal element in S. Then λ is an extension of σ, and we contend that $K = E$. Otherwise, there exists $\alpha \in E$,

$\alpha \notin K$. By what we saw above, our embedding λ has an extension to $K(\alpha)$, thereby contradicting the maximality of (K, λ). This proves that there exists an extension of σ to E. We denote this extension again by σ.

If E is algebraically closed, and L is algebraic over σk, then σE is algebraically closed and L is algebraic over σE, hence $L = \sigma E$.

As a corollary, we have a certain uniqueness for an "algebraic closure" of a field k.

Corollary 2.9. *Let k be a field and let E, E' be algebraic extensions of k. Assume that E, E' are algebraically closed. Then there exists an isomorphism*

$$\tau: E \to E'$$

of E onto E' inducing the identity on k.

Proof. Extend the identity mapping on k to an embedding of E into E' and apply the theorem.

We see that an algebraically closed and algebraic extension of k is determined up to an isomorphism. Such an extension will be called an **algebraic closure** of k, and we frequently denote it by k^a. In fact, unless otherwise specified, we use the symbol k^a only to denote algebraic closure.

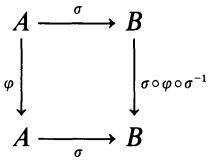
It is now worth while to recall the general situation of isomorphisms and automorphisms in general categories.

Let $\mathfrak{C}$ be a category, and A, B objects in $\mathfrak{C}$. We denote by $\mathrm{Iso}(A, B)$ the set of isomorphisms of A on B. Suppose there exists at least one such isomorphism $\sigma: A \to B$, with inverse $\sigma^{-1}: B \to A$. If φ is an automorphism of A, then $\sigma \circ \varphi: A \to B$ is again an isomorphism. If ψ is an automorphism of B, then $\psi \circ \sigma: A \to B$ is again an isomorphism. Furthermore, the groups of automorphisms $\mathrm{Aut}(A)$ and $\mathrm{Aut}(B)$ are isomorphic, under the mappings

$$\varphi \mapsto \sigma \circ \varphi \circ \sigma^{-1},$$

$$\sigma^{-1} \circ \psi \circ \sigma \hookleftarrow \psi,$$

which are inverse to each other. The isomorphism $\sigma \circ \varphi \circ \sigma^{-1}$ is the one which makes the following diagram commutative:



We have a similar diagram for $\sigma^{-1} \circ \psi \circ \sigma$.

Let $\tau: A \to B$ be another isomorphism. Then $\tau^{-1} \circ \sigma$ is an automorphism of A, and $\tau \circ \sigma^{-1}$ is an automorphism of B. Thus two isomorphisms differ by an automorphism (of A or B). We see that the group $\mathrm{Aut}(B)$ operates on the

set $\mathrm{Iso}(A, B)$ on the left, and $\mathrm{Aut}(A)$ operates on the set $\mathrm{Iso}(A, B)$ on the right.

We also see that $\mathrm{Aut}(A)$ is determined up to a mapping analogous to a conjugation. This is quite different from the type of uniqueness given by universal objects in a category. Such objects have only the identity automorphism, and hence are determined up to a unique isomorphism.

This is not the case with the algebraic closure of a field, which usually has a large amount of automorphisms. Most of this chapter and the next is devoted to the study of such automorphisms.

Examples. It will be proved later in this book that the complex numbers are algebraically closed. Complex conjugation is an automorphism of $\mathbf{C}$. There are many more automorphisms, but the other automorphisms $\neq \mathrm{id}$. are not continuous. We shall discuss other possible automorphisms in the chapter on transcendental extensions. The subfield of $\mathbf{C}$ consisting of all numbers which are algebraic over $\mathbf{Q}$ is an algebraic closure $\mathbf{Q}^a$ of $\mathbf{Q}$. It is easy to see that $\mathbf{Q}^a$ is denumerable. In fact, prove the following as an exercise:

If k is a field which is not finite, then any algebraic extension of k has the same cardinality as k.

If k is denumerable, one can first enumerate all polynomials in k, then enumerate finite extensions by their degree, and finally enumerate the cardinality of an arbitrary algebraic extension. We leave the counting details as exercises.

In particular, $\mathbf{Q}^a \neq \mathbf{C}$. If $\mathbf{R}$ is the field of real numbers, then $\mathbf{R}^a = \mathbf{C}$.

If k is a finite field, then algebraic closure k^a of k is denumerable. We shall in fact describe in great detail the nature of algebraic extensions of finite fields later in this chapter.

Not all interesting fields are subfields of the complex numbers. For instance, one wants to investigate the algebraic extensions of a field $\mathbf{C}(X)$ where X is a variable over $\mathbf{C}$. The study of these extensions amounts to the study of ramified coverings of the sphere (viewed as a Riemann surface), and in fact one has precise information concerning the nature of such extensions, because one knows the fundamental group of the sphere from which a finite number of points has been deleted. We shall mention this example again later when we discuss Galois groups.

§3. SPLITTING FIELDS AND NORMAL EXTENSIONS

Let k be a field and let f be a polynomial in $k[X]$ of degree $\geqq 1$. By a **splitting field** K of f we shall mean an extension K of k such that f splits into linear factors in K, i.e.

$$f(X) = c(X - \alpha_1) \cdots (X - \alpha_n)$$

with $\alpha_i \in K$, $i = 1, \ldots, n$, and such that $K = k(\alpha_1, \ldots, \alpha_n)$ is generated by all the roots of f.

Theorem 3.1. *Let K be a splitting field of the polynomial $f(X) \in k[X]$. If E is another splitting field of f, then there exists an isomorphism $\sigma: E \to K$ inducing the identity on k. If $k \subset K \subset k^a$, where k^a is an algebraic closure of k, then any embedding of E in k^a inducing the identity on k must be an isomorphism of E onto K.*

Proof. Let K^a be an algebraic closure of K. Then K^a is algebraic over k, hence is an algebraic closure of k. By Theorem 2.8 there exists an embedding

$$\sigma: E \to K^a$$

inducing the identity on k. We have a factorization

$$f(X) = c(X - \beta_1) \cdots (X - \beta_n)$$

with $\beta_i \in E$, $i = 1, \ldots, n$. The leading coefficient c lies in k. We obtain

$$f(X) = f^\sigma(X) = c(X - \sigma\beta_1) \cdots (X - \sigma\beta_n).$$

We have unique factorization in $K^a[X]$. Since f has a factorization

$$f(X) = c(X - \alpha_1) \cdots (X - \alpha_n)$$

in $K[X]$, it follows that $(\sigma\beta_1, \ldots, \sigma\beta_n)$ differs from $(\alpha_1, \ldots, \alpha_n)$ by a permutation. From this we conclude that $\sigma\beta_i \in K$ for $i = 1, \ldots, n$ and hence that $\sigma E \subset K$. But $K = k(\alpha_1, \ldots, \alpha_n) = k(\sigma\beta_1, \ldots, \sigma\beta_n)$, and hence $\sigma E = K$, because

$$E = k(\beta_1, \ldots, \beta_n).$$

This proves our theorem.

We note that a polynomial $f(X) \in k[X]$ always has a splitting field, namely the field generated by its roots in a given algebraic closure k^a of k.

Let I be a set of indices and let $\{f_i\}_{i \in I}$ be a family of polynomials in $k[X]$, of degrees $\geqq 1$. By a **splitting field** for this family we shall mean an extension K of k such that every f_i splits in linear factors in $K[X]$, and K is generated by all the roots of all the polynomials f_i, $i \in I$. In most applications we deal with a finite indexing set I, but it is becoming increasingly important to consider infinite algebraic extensions, and so we shall deal with them fairly systematically. One should also observe that the proofs we shall give for various statements would not be simpler if we restricted ourselves to the finite case.

Let k^a be an algebraic closure of k, and let K_i be a splitting field of f_i in k^a. Then the compositum of the K_i is a splitting field for our family,

since the two conditions defining a splitting field are immediately satisfied. Furthermore Theorem 3.1 extends at once to the infinite case:

Corollary 3.2. *Let K be a splitting field for the family $\{f_i\}_{i \in I}$ and let E be another splitting field. Any embedding of E into K^a inducing the identity on k gives an isomorphism of E onto K.*

Proof. Let the notation be as above. Note that E contains a unique splitting field E_i of f_i and K contains a unique splitting field K_i of f_i. Any embedding σ of E into K^a must map E_i onto K_i by Theorem 3.1, and hence maps E into K. Since K is the compositum of the fields K_i, our map σ must send E onto K and hence induces an isomorphism of E onto K.

Remark. If I is finite, and our polynomials are $f_1, \ldots, f_n$, then a splitting field for them is a splitting field for the single polynomial $f(X) = f_1(X) \cdots f_n(X)$ obtained by taking the product. However, even when dealing with finite extensions only, it is convenient to deal simultaneously with sets of polynomials rather than a single one.

Theorem 3.3. *Let K be an algebraic extension of k, contained in an algebraic closure k^a of k. Then the following conditions are equivalent:*

NOR 1. *Every embedding of K in k^a over k induces an automorphism of K.*

NOR 2. *K is the splitting field of a family of polynomials in $k[X]$.*

NOR 3. *Every irreducible polynomial of $k[X]$ which has a root in K splits into linear factors in K.*

Proof. Assume **NOR 1**. Let α be an element of K and let $p_\alpha(X)$ be its irreducible polynomial over k. Let β be a root of p_α in k^a. There exists an isomorphism of $k(\alpha)$ on $k(\beta)$ over k, mapping α on β. Extend this isomorphism to an embedding of K in k^a. This extension is an automorphism σ of K by hypothesis, hence $\sigma\alpha = \beta$ lies in K. Hence every root of p_α lies in K, and p_α splits in linear factors in $K[X]$. Hence K is the splitting field of the family $\{p_\alpha\}_{\alpha \in K}$ as α ranges over all elements of K, and **NOR 2** is satisfied.

Conversely, assume **NOR 2**, and let $\{f_i\}_{i \in I}$ be the family of polynomials of which K is the splitting field. If α is a root of some f_i in K, then for any embedding σ of K in k^a over k we know that $\sigma\alpha$ is a root of f_i. Since K is generated by the roots of all the polynomials f_i, it follows that σ maps K into itself. We now apply Lemma 2.1 to conclude that σ is an automorphism.

Our proof that **NOR 1** implies **NOR 2** also shows that **NOR 3** is satisfied. Conversely, assume **NOR 3**. Let σ be an embedding of K in k^a over k. Let $\alpha \in K$ and let $p(X)$ be its irreducible polynomial over k. If σ is an embedding of K in k^a over k then σ maps α on a root β of $p(X)$, and by hypothesis β lies in K. Hence $\sigma\alpha$ lies in K, and σ maps K into itself. By Lemma 2.1, it follows that σ is an automorphism.

An extension K of k satisfying the hypotheses **NOR 1**, **NOR 2**, **NOR 3** will be said to be **normal**. It is not true that the class of normal extensions is distinguished. For instance, it is easily shown that an extension of degree 2 is normal, but the extension $\mathbf{Q}(\sqrt[4]{2})$ of the rational numbers is not normal (the complex roots of $X^4 - 2$ are not in it), and yet this extension is obtained by successive extensions of degree 2, namely

$$E = \mathbf{Q}(\sqrt[4]{2}) \supset F \supset \mathbf{Q},$$

where

$$F = \mathbf{Q}(\alpha), \qquad \alpha = \sqrt{2} \quad \text{and} \quad E = F(\sqrt{\alpha}).$$

Thus a tower of normal extensions is not necessarily normal. However, we still have some of the properties:

Theorem 3.4. *Normal extensions remain normal under lifting. If $K \supset E \supset k$ and K is normal over k, then K is normal over E. If K_1, K_2 are normal over k and are contained in some field L, then K_1K_2 is normal over k, and so is $K_1 \cap K_2$.*

Proof. For our first assertion, let K be normal over k, let F be any extension of k, and assume K, F are contained in some bigger field. Let σ be an embedding of KF over F (in F^a). Then σ induces the identity on F, hence on k, and by hypothesis its restriction to K maps K into itself. We get $(KF)^\sigma = K^\sigma F^\sigma = KF$ whence KF is normal over F.

Assume that $K \supset E \supset k$ and that K is normal over k. Let σ be an embedding of K over E. Then σ is also an embedding of K over k, and our assertion follows by definition.

Finally, if K_1, K_2 are normal over k, then for any embedding σ of K_1K_2 over k we have

$$\sigma(K_1K_2) = \sigma(K_1)\sigma(K_2)$$

and our assertion again follows from the hypothesis. The assertion concerning the intersection is true because

$$\sigma(K_1 \cap K_2) = \sigma(K_1) \cap \sigma(K_2).$$

We observe that if K is a finitely generated normal extension of k, say

$$K = k(\alpha_1, \ldots, \alpha_n),$$

and $p_1, \ldots, p_n$ are the respective irreducible polynomials of $\alpha_1, \ldots, \alpha_n$ over k then K is already the splitting field of the finite family $p_1, \ldots, p_n$. We shall investigate later when K is the splitting field of a single irreducible polynomial.

§4. SEPARABLE EXTENSIONS

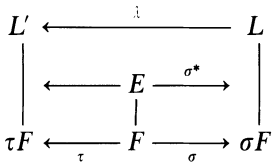
Let E be an algebraic extension of a field F and let

$$\sigma: F \to L$$

be an embedding of F in an algebraically closed field L. We investigate more closely extensions of σ to E. Any such extension of σ maps E on a subfield of L which is algebraic over σF. Hence for our purposes, we shall assume that L is algebraic over σF, hence is equal to an algebraic closure of σF.

Let S_σ be the set of extensions of σ to an embedding of E in L.

Let L' be another algebraically closed field, and let $\tau: F \to L'$ be an embedding. We assume as before that L' is an algebraic closure of τF. By Theorem 2.8, there exists an isomorphism $\lambda: L \to L'$ extending the map $\tau \circ \sigma^{-1}$ applied to the field σF. This is illustrated in the following diagram:



We let S_τ be the set of embeddings of E in L' extending τ.

If $\sigma^* \in S_\sigma$ is an extension of σ to an embedding of E in L, then $\lambda \circ \sigma^*$ is an extension of τ to an embedding of E into L', because for the restriction to F we have

$$\lambda \circ \sigma^* = \tau \circ \sigma^{-1} \circ \sigma = \tau.$$

Thus λ induces a mapping from S_σ into S_τ. It is clear that the inverse mapping is induced by λ^{-1}, and hence that S_σ, S_τ are in bijection under the mapping

$$\sigma^* \mapsto \lambda \circ \sigma^*.$$

In particular, the cardinality of S_σ, S_τ is the same. Thus this cardinality depends only on the extension E/F, and will be denoted by

$$[E:F]_s.$$

We shall call it the **separable degree** of E over F. It is mostly interesting when E/F is finite.

Theorem 4.1. *Let $E \supset F \supset k$ be a tower. Then*

$$[E:k]_s = [E:F]_s[F:k]_s.$$

Furthermore, if E is finite over k, then $[E:k]_s$ is finite and

$$[E:k]_s \leqq [E:k].$$

The separable degree is at most equal to the degree.

Proof. Let $\sigma: k \to L$ be an embedding of k in an algebraically closed field L. Let $\{\sigma_i\}_{i \in I}$ be the family of distinct extensions of σ to F, and for each i, let $\{\tau_{ij}\}$ be the family of distinct extensions of σ_i to E. By what we saw before, each σ_i has precisely $[E:F]_s$ extensions to embeddings of E in L. The set of embeddings $\{\tau_{ij}\}$ contains precisely

$$[E:F]_s[F:k]_s$$

elements. Any embedding of E into L over σ must be one of the τ_{ij}, and thus we see that the first formula holds, i.e. we have multiplicativity in towers.

As to the second, let us assume that E/k is finite. Then we can obtain E as a tower of extensions, each step being generated by one element:

$$k \subset k(\alpha_1) \subset k(\alpha_1, \alpha_2) \subset \cdots \subset k(\alpha_1, \ldots, \alpha_r) = E.$$

If we define inductively $F_{v+1} = F_v(\alpha_{v+1})$ then by Proposition 2.7,

$$[F_v(\alpha_{v+1}):F_v]_s \leqq [F_v(\alpha_{v+1}):F_v].$$

Thus our inequality is true in each step of the tower. By multiplicativity, it follows that the inequality is true for the extension E/k, as was to be shown.

Corollary 4.2. *Let E be finite over k, and $E \supset F \supset k$. The equality*

$$[E:k]_s = [E:k]$$

holds if and only if the corresponding equality holds in each step of the tower, i.e. for E/F and F/k.

Proof. Clear.

It will be shown later (and it is not difficult to show) that $[E:k]_s$ divides the degree $[E:k]$ when E is finite over k. We define $[E:k]_i$ to be the quotient, so that

$$[E:k]_s[E:k]_i = [E:k].$$

It then follows from the multiplicativity of the separable degree and of the degree in towers that the symbol $[E:k]_i$ is also multiplicative in towers. We shall deal with it at greater length in §6.

Let E be a finite extension of k. We shall say that E is **separable** over k if $[E:k]_s = [E:k]$.

An element α algebraic over k is said to be **separable** over k if $k(\alpha)$ is separable over k. We see that this condition is equivalent to saying that the irreducible polynomial $\mathrm{Irr}(\alpha, k, X)$ has no multiple roots.

A polynomial $f(X) \in k[X]$ is called **separable** if it has no multiple roots.

If α is a root of a separable polynomial $g(X) \in k[X]$ then the irreducible polynomial of α over k divides g and hence α is separable over k.

We note that if $k \subset F \subset K$ and $\alpha \in K$ is separable over k, then α is separable over F. Indeed, if f is a separable polynomial in $k[X]$ such that $f(\alpha) = 0$, then f also has coefficients in F, and thus α is separable over F. (We may say that a separable element remains separable under lifting.)

Theorem 4.3. *Let E be a finite extension of k. Then E is separable over k if and only if each element of E is separable over k.*

Proof. Assume E is separable over k and let $\alpha \in E$. We consider the tower

$$k \subset k(\alpha) \subset E.$$

By Corollary 4.2, we must have $[k(\alpha):k] = [k(\alpha):k]_s$ whence α is separable over k. Conversely, assume that each element of E is separable over k. We can write $E = k(\alpha_1, \ldots, \alpha_n)$ where each α_i is separable over k. We consider the tower

$$k \subset k(\alpha_1) \subset k(\alpha_1, \alpha_2) \subset \cdots \subset k(\alpha_1, \ldots, \alpha_n).$$

Since each α_i is separable over k, each α_i is separable over $k(\alpha_1, \ldots, \alpha_{i-1})$ for $i \geqq 2$. Hence by the tower theorem, it follows that E is separable over k.

We observe that our last argument shows: If E is generated by a finite number of elements, each of which is separable over k, then E is separable over k.

Let E be an arbitrary algebraic extension of k. We define E to be **separable** over k if every finitely generated subextension is separable over k, i.e., if every extension $k(\alpha_1, \ldots, \alpha_n)$ with $\alpha_1, \ldots, \alpha_n \in E$ is separable over k.

Theorem 4.4. *Let E be an algebraic extension of k, generated by a family of elements $\{\alpha_i\}_{i \in I}$. If each α_i is separable over k then E is separable over k.*

Proof. Every element of E lies in some finitely generated subfield

$$k(\alpha_{i_1}, \ldots, a_{i_n}),$$

and as we remarked above, each such subfield is separable over k. Hence every element of E is separable over k by Theorem 4.3, and this concludes the proof.

Theorem 4.5. *Separable extensions form a distinguished class of extensions.*

Proof. Assume that E is separable over k and let $E \supset F \supset k$. Every element of E is separable over F, and every element of F is an element of E, so separable over k. Hence each step in the tower is separable. Conversely, assume that $E \supset F \supset k$ is some extension such that E/F is separable and F/k is separable. If E is finite over k, then we can use Corollary 4.2. Namely, we have an equality of the separable degree and the degree in each step of the tower, whence an equality for E over k by multiplicativity.

If E is infinite, let $\alpha \in E$. Then α is a root of a separable polynomial $f(X)$ with coefficients in F. Let these coefficients be $a_n, \ldots, a_0$. Let $F_0 = k(a_n, \ldots, a_0)$. Then F_0 is separable over k, and α is separable over F_0. We now deal with the finite tower

$$k \subset F_0 \subset F_0(\alpha)$$

and we therefore conclude that $F_0(\alpha)$ is separable over k, hence that α is separable over k. This proves condition (1) in the definition of "distinguished."

Let E be separable over k. Let F be any extension of k, and assume that E, F are both subfields of some field. Every element of E is separable over k, whence separable over F. Since EF is generated over F by all the elements of E, it follows that EF is separable over F, by Theorem 4.4. This proves condition (2) in the definition of "distinguished," and concludes the proof of our theorem.

Let E be a finite extension of k. The intersection of all normal extensions K of k (in an algebraic closure E^a) containing E is a normal extension of k which contains E, and is obviously the smallest normal extension of k containing E. If $\sigma_1, \ldots, \sigma_n$ are the distinct embeddings of E in E^a, then the extension

$$K = (\sigma_1 E)(\sigma_2 E) \cdots (\sigma_n E),$$

which is the compositum of all these embeddings, is a normal extension of k, because for any embedding of it, say τ, we can apply τ to each extension $\sigma_i E$. Then $(\tau\sigma_1, \ldots, \tau\sigma_n)$ is a permutation of $(\sigma_1, \ldots, \sigma_n)$ and thus τ maps K into itself. Any normal extension of k containing E must contain $\sigma_i E$ for each i, and thus *the smallest normal extension of k containing E is precisely equal to the compositum*

$$(\sigma_1 E) \cdots (\sigma_n E).$$

If E is separable over k, then from Theorem 4.5 and induction we conclude that the smallest normal extension of k containing E is also separable over k.

Similar results hold for an infinite algebraic extension E of k, taking an infinite compositum.

In light of Theorem 4.5, the compositum of all separable extensions of a field k in a given algebraic closure k^a is a separable extension, which will be denoted by k^s or k^{sep}, and will be called the **separable closure** of k. As a matter of terminology, if E is an algebraic extension of k, and σ any embedding of E in k^a over k, then we call σE a **conjugate** of E in k^a. *We can say that the smallest normal extension of k containing E is the compositum of all the conjugates of E in E^a.*

Let α be algebraic over k. If $\sigma_1, \ldots, \sigma_r$ are the distinct embeddings of $k(\alpha)$ into k^a over k, then we call $\sigma_1\alpha, \ldots, \sigma_r\alpha$ the **conjugates** of α in k^a. These elements are simply the distinct roots of the irreducible polynomial of α over k. The smallest normal extension of k containing one of these conjugates is simply $k(\sigma_1\alpha, \ldots, \sigma_r\alpha)$.

Theorem 4.6. (Primitive Element Theorem). *Let E be a finite extension of a field k. There exists an element $\alpha \in E$ such that $E = k(\alpha)$ if and only if there exists only a finite number of fields F such that $k \subset F \subset E$. If E is separable over k, then there exists such an element α.*

Proof. If k is finite, then we know that the multiplicative group of E is generated by one element, which will therefore also generate E over k. We assume that k is infinite.

Assume that there is only a finite number of fields, intermediate between k and E. Let $\alpha, \beta \in E$. As c ranges over elements of k, we can only have a finite number of fields of type $k(\alpha + c\beta)$. Hence there exist elements c_1, $c_2 \in k$ with $c_1 \neq c_2$ such that

$$k(\alpha + c_1\beta) = k(\alpha + c_2\beta).$$

Note that $\alpha + c_1\beta$ and $\alpha + c_2\beta$ are in the same field, whence so is $(c_1 - c_2)\beta$, and hence so is β. Thus α is also in that field, and we see that $k(\alpha, \beta)$ can be generated by one element.

Proceeding inductively, if $E = k(\alpha_1, \ldots, \alpha_n)$ then there will exist elements $c_2, \ldots, c_n \in k$ such that

$$E = k(\xi)$$

where $\xi = \alpha_1 + c_2\alpha_2 + \cdots + c_n\alpha_n$. This proves half of our theorem.

Conversely, assume that $E = k(\alpha)$ for some α, and let $f(X) = \mathrm{Irr}(\alpha, k, X)$. Let $k \subset F \subset E$. Let $g_F(X) = \mathrm{Irr}(\alpha, F, X)$. Then g_F divides f. We have unique factorization in $E[X]$, and any polynomial in $E[X]$ which has leading coefficient 1 and divides $f(X)$ is equal to a product of factors $(X - \alpha_i)$ where $\alpha_1, \ldots, \alpha_n$ are the roots of f in a fixed algebraic closure. Hence there is only a finite number of such polynomials. Thus we get a mapping

$$F \mapsto g_F$$

from the set of intermediate fields into a finite set of polynomials. Let F_0 be

the subfield of F generated over k by the coefficients of $g_F(X)$. Then g_F has coefficients in F_0 and is irreducible over F_0 since it is irreducible over F. Hence the degree of α over F_0 is the same as the degree of α over F. Hence $F = F_0$. Thus our field F is uniquely determined by its associated polynomials g_F, and our mapping is therefore injective. This proves the first assertion of the theorem.

As to the statement concerning separable extensions, using induction, we may assume without loss of generality that $E = k(\alpha, \beta)$ where α, β are separable over k. Let $\sigma_1, \ldots, \sigma_n$ be the distinct embeddings of $k(\alpha, \beta)$ in k^a over k. Let

$$P(X) = \prod_{i \neq j} (\sigma_i\alpha + X\sigma_i\beta - \sigma_j\alpha - X\sigma_j\beta).$$

Then $P(X)$ is not the zero polynomial, and hence there exists $c \in k$ such that $P(c) \neq 0$. Then the elements $\sigma_i(\alpha + c\beta)$ $(i = 1, \ldots, n)$ are distinct, whence $k(\alpha + c\beta)$ has degree at least n over k. But $n = [k(\alpha, \beta) : k]$, and hence

$$k(\alpha, \beta) = k(\alpha + c\beta),$$

as desired.

If $E = k(\alpha)$, then we say that α is a **primitive element** of E (over k).

§5. FINITE FIELDS

We have developed enough general theorems to describe the structure of finite fields. This is interesting for its own sake, and also gives us examples for the general theory.

Let F be a finite field with q elements. As we have noted previously, we have a homomorphism

$$\mathbf{Z} \to F$$

sending 1 on 1, whose kernel cannot be 0, and hence is a principal ideal generated by a prime number p since $\mathbf{Z}/p\mathbf{Z}$ is embedded in F and F has no divisors of zero. Thus F has characteristic p, and contains a field isomorphic to $\mathbf{Z}/p\mathbf{Z}$.

We remark that $\mathbf{Z}/p\mathbf{Z}$ has no automorphisms other than the identity. Indeed, any automorphism must map 1 on 1, hence leaves every element fixed because 1 generates $\mathbf{Z}/p\mathbf{Z}$ additively. We identify $\mathbf{Z}/p\mathbf{Z}$ with its image in F. Then F is a vector space over $\mathbf{Z}/p\mathbf{Z}$, and this vector space must be

finite since F is finite. Let its degree be n. Let $\omega_1, \ldots, \omega_n$ be a basis for F over $\mathbf{Z}/p\mathbf{Z}$. Every element of F has a unique expression of the form

$$a_1\omega_1 + \cdots + a_n\omega_n$$

with $a_i \in \mathbf{Z}/p\mathbf{Z}$. Hence $q = p^n$.

The multiplicative group F^* of F has order $q - 1$. Every $\alpha \in F^*$ satisfies the equation $X^{q-1} = 1$. Hence every element of F satisfies the equation

$$f(X) = X^q - X = 0.$$

This implies that the polynomial $f(X)$ has q distinct roots in F, namely all elements of F. Hence f splits into factors of degree 1 in F, namely

$$X^q - X = \prod_{\alpha \in F} (X - \alpha).$$

In particular, F is a splitting field for f. But a splitting field is uniquely determined up to an isomorphism. Hence if a finite field of order p^n exists, it is uniquely determined, up to an isomorphism, as the splitting field of $X^{p^n} - X$ over $\mathbf{Z}/p\mathbf{Z}$.

As a matter of notation, we denote $\mathbf{Z}/p\mathbf{Z}$ by $\mathbf{F}_p$. Let n be an integer $\geqq 1$ and consider the splitting field of

$$X^{p^n} - X = f(X)$$

in an algebraic closure $\mathbf{F}_p^a$. We contend that this splitting field is the set of roots of $f(X)$ in $\mathbf{F}_p^a$. Indeed, let α, β be roots. Then

$$(\alpha + \beta)^{p^n} - (\alpha + \beta) = \alpha^{p^n} + \beta^{p^n} - \alpha - \beta = 0,$$

whence $\alpha + \beta$ is a root. Also,

$$(\alpha\beta)^{p^n} - \alpha\beta = \alpha^{p^n}\beta^{p^n} - \alpha\beta = \alpha\beta - \alpha\beta = 0,$$

and $\alpha\beta$ is a root. Note that 0, 1 are roots of $f(X)$. If $\beta \neq 0$ then

$$(\beta^{-1})^{p^n} - \beta^{-1} = (\beta^{p^n})^{-1} - \beta^{-1} = 0$$

so that β^{-1} is a root. Finally,

$$(-\beta)^{p^n} - (-\beta) = (-1)^{p^n}\beta^{p^n} + \beta.$$

If p is odd, then $(-1)^{p^n} = -1$ and we see that $-\beta$ is a root. If p is even then $-1 = 1$ (in $\mathbf{Z}/2\mathbf{Z}$) and hence $-\beta = \beta$ is a root. This proves our contention.

The derivative of $f(X)$ is

$$f'(X) = p^n X^{p^n - 1} - 1 = -1.$$

Hence $f(X)$ has no multiple roots, and therefore has p^n distinct roots in $\mathbf{F}_p^a$. Hence its splitting field has exactly p^n elements. We summarize our results:

Theorem 5.1. *For each prime p and each integer $n \geqq 1$ there exists a finite field of order p^n denoted by $\mathbf{F}_{p^n}$, uniquely determined as a subfield of an algebraic closure $\mathbf{F}_p^a$. It is the splitting field of the polynomial*

$$X^{p^n} - X,$$

and its elements are the roots of this polynomial. Every finite field is isomorphic to exactly one field $\mathbf{F}_{p^n}$.

We usually write $p^n = q$ and $\mathbf{F}_q$ instead of $\mathbf{F}_{p^n}$.

Corollary 5.2. *Let $\mathbf{F}_q$ be a finite field. Let n be an integer $\geqq 1$. In a given algebraic closure $\mathbf{F}_q^a$, there exists one and only one extension of $\mathbf{F}_q$ of degree n, and this extension is the field $\mathbf{F}_{q^n}$.*

Proof. Let $q = p^m$. Then $q^n = p^{mn}$. The splitting field of $X^{q^n} - X$ is precisely $F_{p^{mn}}$ and has degree mn over $\mathbf{Z}/p\mathbf{Z}$. Since $\mathbf{F}_q$ has degree m over $\mathbf{Z}/p\mathbf{Z}$, it follows that $\mathbf{F}_{q^n}$ has degree n over $\mathbf{F}_q$. Conversely, any extension of degree n over $\mathbf{F}_q$ has degree mn over $\mathbf{F}_p$ and hence must be $\mathbf{F}_{p^{mn}}$. This proves our corollary.

Theorem 5.3. *The multiplicative group of a finite field is cyclic.*

Proof. This has already been proved in Chapter IV, Theorem 1.9.

We shall determine all automorphisms of a finite field.

Let $q = p^n$ and let $\mathbf{F}_q$ be the finite field with q elements. We consider the **Frobenius mapping**

$$\varphi: \mathbf{F}_q \to \mathbf{F}_q$$

such that $\varphi(x) = x^p$. Then φ is a homomorphism, and its kernel is 0 since $\mathbf{F}_q$ is a field. Hence φ is injective. Since $\mathbf{F}_q$ is finite, it follows that φ is surjective, and hence that φ is an isomorphism. We note that it leaves $\mathbf{F}_p$ fixed.

Theorem 5.4. *The group of automorphisms of $\mathbf{F}_q$ is cyclic of degree n, generated by φ.*

Proof. Let G be the group generated by φ. We note that $\varphi^n = \mathrm{id}$ because $\varphi^n(x) = x^{p^n} = x$ for all $x \in \mathbf{F}_q$. Hence n is an exponent for φ. Let d be the period of φ, so $d \geqq 1$. We have $\varphi^d(x) = x^{p^d}$ for all $x \in \mathbf{F}_q$. Hence each $x \in \mathbf{F}_q$ is a root of the equation

$$X^{p^d} - X = 0.$$

This equation has at most p^d roots. It follows that $d \geqq n$, whence $d = n$.

There remains to be proved that G is the group of all automorphisms of $\mathbf{F}_q$. Any automorphism of $\mathbf{F}_q$ must leave $\mathbf{F}_p$ fixed. Hence it is an auto-

morphism of $\mathbf{F}_q$ over $\mathbf{F}_p$. By Theorem 4.1, the number of such automorphisms is $\leqq n$. Hence $\mathbf{F}_q$ cannot have any other automorphisms except for those of G.

Theorem 5.5. *Let m, n be integers $\geqq 1$. Then in any algebraic closure of $\mathbf{F}_p$, the subfield $\mathbf{F}_{p^n}$ is contained in $\mathbf{F}_{p^m}$ if and only if n divides m. If that is the case, let $q = p^n$, and let $m = nd$. Then $\mathbf{F}_{p^m}$ is normal and separable over $\mathbf{F}_q$, and the group of automorphisms of $\mathbf{F}_{p^m}$ over $\mathbf{F}_q$ is cyclic of order d, generated by φ^n.*

Proof. All the statements are trivial consequences of what has already been proved and will be left to the reader.

§6. INSEPARABLE EXTENSIONS

This section is of a fairly technical nature, and can be omitted without impairing the understanding of most of the rest of the book.

We begin with some remarks supplementing those of Proposition 2.7.

Let $f(X) = (X - \alpha)^m g(X)$ be a polynomial in $k[X]$, and assume $X - \alpha$ does not divide $g(X)$. We recall that m is called the multiplicity of α in f. We say that α is a **multiple** root of f if $m > 1$. Otherwise, we say that α is a **simple** root.

Proposition 6.1. *Let α be algebraic over k, $\alpha \in k^a$, and let*

$$f(X) = \operatorname{Irr}(\alpha, k, X).$$

If $\operatorname{char} k = 0$, *then all roots of f have multiplicity* 1 (*f is separable*). *If*

$$\operatorname{char} k = p > 0,$$

then there exists an integer $\mu \geqq 0$ such that every root of f has multiplicity p^μ. We have

$$[k(\alpha) : k] = p^\mu [k(\alpha) : k]_s,$$

and α^{p^μ} is separable over k.

Proof. Let $\alpha_1, \ldots, \alpha_r$ be the distinct roots of f in k^a and let $\alpha = \alpha_1$. Let m be the multiplicity of α in f. Given $1 \leqq i \leqq r$, there exists an isomorphism

$$\sigma : k(\alpha) \to k(\alpha_i)$$

over k such that $\sigma\alpha = \alpha_i$. Extend σ to an automorphism of k^a and denote

this extension also by σ. Since f has coefficients in k we have $f^\sigma = f$. We note that

$$f(X) = \prod_{j=1}^{r} (X - \sigma\alpha_j)^{m_j}$$

if m_j is the multiplicity of α_j in f. By unique factorization, we conclude that $m_i = m_1$ and hence that all m_i are equal to the same integer m.

Consider the derivative $f'(X)$. If f and f' have a root in common, then α is a root of a polynomial of lower degree than $\deg f$. This is impossible unless $\deg f' = -\infty$, in other words, f' is identically 0. If the characteristic is 0, this cannot happen. Hence if f has multiple roots, we are in characteristic p, and $f(X) = g(X^p)$ for some polynomial $g(X) \in k[X]$. Therefore α^p is a root of a polynomial g whose degree is $< \deg f$. Proceeding inductively, we take the smallest integer $\mu \geqq 0$ such that α^{p^μ} is the root of a separable polynomial in $k[X]$, namely the polynomial h such that

$$f(X) = h(X^{p^\mu}).$$

Comparing the degree of f and g, we conclude that

$$[k(\alpha) : k(\alpha^p)] = p.$$

Inductively, we find

$$[k(\alpha) : k(\alpha^{p^\mu})] = p^\mu.$$

Since h has roots of multiplicity 1, we know that

$$[k(\alpha^{p^\mu}) : k]_s = [k(\alpha^{p^\mu}) : k],$$

and comparing the degree of f and the degree of h, we see that the number of distinct roots of f is equal to the number of distinct roots of h. Hence

$$[k(\alpha) : k]_s = [k(\alpha^{p^\mu}) : k]_s.$$

From this our formula for the degree follows by multiplicativity, and our proposition is proved. We note that the roots of h are

$$\alpha_1^{p^\mu}, \ldots, \alpha_r^{p^\mu}.$$

Corollary 6.2. *For any finite extension E of k, the separable degree $[E : k]_s$ divides the degree $[E : k]$. The quotient is 1 if the characteristic is 0, and a power of p if the characteristic is $p > 0$.*

Proof. We decompose E/k into a tower, each step being generated by one element, and apply Proposition 6.1, together with the multiplicativity of our indices in towers.

If E/K is finite, we call the quotient

$$\frac{[E:k]}{[E:k]_s}$$

the **inseparable degree** (or **degree of inseparability**), and denote it by $[E:k]_i$ as in §4. We have

$$[E:k]_s[E:k]_i = [E:k].$$

Corollary 6.3. *A finite extension is separable if and only if* $[E:k]_i = 1$.

Proof. By definition.

Corollary 6.4 *If* $E \supset F \supset k$ *are two finite extensions, then*

$$[E:k]_i = [E:F]_i[F:k]_i.$$

Proof. Immediate by Theorem 4.1.

We now assume throughout that k is a field of characteristic $p > 0$.

An element α algebraic over k is said to be **purely inseparable** over k if there exists an integer $n \geqq 0$ such that α^{p^n} lies in k.

Let E be an algebraic extension of k. We contend that the following conditions are equivalent:

P. Ins. 1. We have $[E:k]_s = 1$.

P. Ins. 2. Every element α of E is purely inseparable over k.

P. Ins. 3. For every $\alpha \in E$, the irreducible equation of α over k is of type $X^{p^n} - a = 0$ with some $n \geqq 0$ and $a \in k$.

P. Ins. 4. There exists a set of generators $\{\alpha_i\}_{i \in I}$ of E over k such that each α_i is purely inseparable over k.

To prove the equivalence, assume **P. Ins. 1**. Let $\alpha \in E$. By Theorem 4.1, we conclude that $[k(\alpha):k]_s = 1$. Let $f(X) = \mathrm{Irr}(\alpha, k, X)$. Then f has only one root since

$$[k(\alpha):k]_s$$

is equal to the number of distinct roots of $f(X)$. Let $m = [k(\alpha):k]$. Then $\deg f = m$, and the factorization of f over $k(\alpha)$ is $f(X) = (X - \alpha)^m$. Write $m = p^n r$ where r is an integer prime to p. Then

$$\begin{aligned} f(X) &= (X^{p^n} - \alpha^{p^n})^r \\ &= X^{p^n r} - r\alpha^{p^n} X^{p^n(r-1)} + \text{lower terms.} \end{aligned}$$

Since the coefficients of $f(X)$ lie in k, it follows that

$$r\alpha^{p^n}$$

lies in k, and since $r \neq 0$ (in k), then α^{p^n} lies in k. Let $a = \alpha^{p^n}$. Then α is a root of the polynomial $X^{p^n} - a$, which divides $f(X)$. It follows that $f(X) = X^{p^n} - a$.

Essentially the same argument as the preceding one shows that **P. Ins. 2** implies **P. Ins. 3**. It is trivial that the third condition implies the fourth.

Finally, assume **P. Ins. 4**. Let E be an extension generated by purely inseparable elements α_i ($i \in I$). Any embedding of E over k maps α_i on a root of

$$f_i(X) = \operatorname{Irr}(\alpha_i, k, X).$$

But $f_i(X)$ divides some polynomial $X^{p^n} - a$, which has only one root. Hence any embedding of E over k is the identity on each α_i, whence the identity on E, and we conclude that $[E : k]_s = 1$, as desired.

An extension satisfying the above four properties will be called **purely inseparable**.

Proposition 6.5. *Purely inseparable extensions form a distinguished class of extensions.*

Proof. The tower theorem is clear from Theorem 4.1, and the lifting property is clear from condition **P. Ins. 4**.

Proposition 6.6. *Let E be an algebraic extension of k. Let E_0 be the compositum of all subfields F of E such that $F \supset k$ and F is separable over k. Then E_0 is separable over k, and E is purely inseparable over E_0.*

Proof. Since separable extensions form a distinguished class, we know that E_0 is separable over k. In fact, E_0 consists of all elements of E which are separable over k. By Proposition 6.1, given $\alpha \in E$ there exists a power of p, say p^n such that α^{p^n} is separable over k. Hence E is purely inseparable over E_0, as was to be shown.

Corollary 6.7. *If an algebraic extension E of k is both separable and purely inseparable, then $E = k$.*

Proof. Obvious.

Corollary 6.8. *Let K be normal over k and let K_0 be its maximal separable subextension. Then K_0 is also normal over k.*

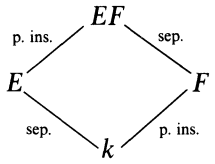
Proof. Let σ be an embedding of K_0 in K^a over k and extend σ to an embedding of K. Then σ is an automorphism of K. Furthermore, σK_0 is separable over k, hence is contained in K_0, since K_0 is the maximal separable subfield. Hence $\sigma K_0 = K_0$, as contended.

Corollary 6.9. *Let E, F be two finite extensions of k, and assume that E/k is separable, F/k is purely inseparable. Assume E, F are subfields of a common field. Then*

$$[EF:F] = [E:k] = [EF:k]_s,$$

$$[EF:E] = [F:k] = [EF:k]_i.$$

Proof. The picture is as follows:



The proof is a trivial juggling of indices, using the corollaries of Proposition 6.1. We leave it as an exercise.

Corollary 6.10. *Let E^p denote the field of all elements x^p, $x \in E$. Let E be a finite extension of k. If $E^p k = E$, then E is separable over k. If E is separable over k, then $E^{p^n} k = E$ for all $n \geqq 1$.*

Proof. Let E_0 be the maximal separable subfield of E. Assume $E^p k = E$. Let $E = k(\alpha_1, \ldots, \alpha_n)$. Since E is purely inseparable over E_0 there exists m such that $\alpha_i^{p^m} \in E_0$ for each $i = 1, \ldots, n$. Hence $E^{p^m} \subset E_0$. But $E^{p^m} k = E$ whence $E = E_0$ is separable over k. Conversely, assume that E is separable over k. Then E is separable over $E^p k$. Since E is also purely inseparable over $E^p k$ we conclude that $E = E^p k$. Similarly we get $E = E^{p^n} k$ *for* $n \geqq 1$, as was to be shown.

Proposition 6.6 shows that any algebraic extension can be decomposed into a tower consisting of a maximal separable subextension and a purely inseparable step above it. Usually, one cannot reverse the order of the tower. However, there is an important case when it can be done.

Proposition 6.11. *Let K be normal over k. Let G be its group of automorphisms over k. Let K^G be the fixed field of G (see Chapter* VI, §1*). Then K^G is purely inseparable over k, and K is separable over K^G. If K_0 is the maximal separable subextension of K, then $K = K^G K_0$ and $K_0 \cap K^G = k$.*

Proof. Let $\alpha \in K^G$. Let τ be an embedding of $k(\alpha)$ over k in K^a and extend τ to an embedding of K, which we denote also by τ. Then τ is an automorphism of K because K is normal over k. By definition, $\tau\alpha = \alpha$ and hence τ is the identity on $k(\alpha)$. Hence $[k(\alpha):k]_s = 1$ and α is purely inseparable. Thus K^G is purely inseparable over k. The intersection of K_0

and K^G is both separable and purely inseparable over k, and hence is equal to k.

To prove that K is separable over K^G, assume first that K is finite over k, and hence that G is finite, by Theorem 4.1. Let $\alpha \in K$. Let $\sigma_1, \ldots, \sigma_r$ be a maximal subset of elements of G such that the elements

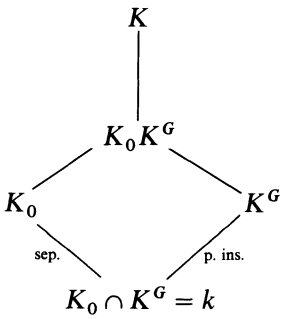
$$\sigma_1\alpha, \ldots, \sigma_r\alpha$$

are distinct, and such that σ_1 is the identity, and α is a root of the polynomial

$$f(X) = \prod_{i=1}^{r} (X - \sigma_i\alpha).$$

For any $\tau \in G$ we note that $f^\tau = f$ because τ permutes the roots. We note that f is separable, and that its coefficients are in the fixed field K^G. Hence α is separable over K^G. The reduction of the infinite case to the finite case is done by observing that every $\alpha \in K$ is contained in some finite normal subextension of K. We leave the details to the reader.

We now have the following picture:



By Proposition 6.6, K is purely inseparable over K_0, hence purely inseparable over K_0K^G. Furthermore, K is separable over K^G, hence separable over K_0K^G. Hence $K = K_0K^G$, thereby proving our proposition.

We see that every normal extension decomposes into a compositum of a purely inseparable and a separable extension. We shall define a Galois extension in the next chapter to be a normal separable extension. Then K_0 is Galois over k and the normal extension is decomposed into a Galois and a purely inseparable extension. The group G is called the **Galois group** of the extension K/k.

A field k is called **perfect** if $k^p = k$. (Every field of characteristic zero is also called perfect.)

Corollary 6.12. *If k is perfect, then every algebraic extension of k is separable, and every algebraic extension of k is perfect.*

Proof. Every finite algebraic extension is contained in a normal extension, and we apply Proposition 6.11 to get what we want.

EXERCISES

1. Let $E = \mathbf{Q}(\alpha)$, where α is a root of the equation

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

Express $(\alpha^2 + \alpha + 1)(\alpha^2 + \alpha)$ and $(\alpha - 1)^{-1}$ in the form

$$a\alpha^2 + b\alpha + c$$

with $a, b, c \in \mathbf{Q}$.

2. Let $E = F(\alpha)$ where α is algebraic over F, of odd degree. Show that $E = F(\alpha^2)$.

3. Let α and β be two elements which are algebraic over F. Let $f(X) = \operatorname{Irr}(\alpha, F, X)$ and $g(X) = \operatorname{Irr}(\beta, F, X)$. Suppose that $\deg f$ and $\deg g$ are relatively prime. Show that g is irreducible in the polynomial ring $F(\alpha)[X]$.

4. Let α be the real positive fourth root of 2. Find all intermediate fields in the extension $\mathbf{Q}(\alpha)$ of $\mathbf{Q}$.

5. If α is a complex root of $X^6 + X^3 + 1$, find all homomorphisms $\sigma: \mathbf{Q}(\alpha) \to \mathbf{C}$. [*Hint*: The polynomial is a factor of $X^9 - 1$.]

6. Show that $\sqrt{2} + \sqrt{3}$ is algebraic over $\mathbf{Q}$, of degree 4.

7. Let E, F be two finite extensions of a field k, contained in a larger field K. Show that

$$[EF : k] \leqq [E : k][F : k].$$

If $[E : k]$ and $[F : k]$ are relatively prime, show that one has an equality sign in the above relation.

8. Let $f(X) \in k[X]$ be a polynomial of degree n. Let K be its splitting field. Show that $[K : k]$ divides $n!$

9. Find the splitting field of $X^{p^8} - 1$ over the field $\mathbf{Z}/p\mathbf{Z}$.

10. Let α be a real number such that $\alpha^4 = 5$.
    (a) Show that $\mathbf{Q}(i\alpha^2)$ is normal over $\mathbf{Q}$.
    (b) Show that $\mathbf{Q}(\alpha + i\alpha)$ is normal over $\mathbf{Q}(i\alpha^2)$.
    (c) Show that $\mathbf{Q}(\alpha + i\alpha)$ is not normal over $\mathbf{Q}$.

11. Describe the splitting fields of the following polynomials over $\mathbf{Q}$, and find the degree of each such splitting field.
    (a) $X^2 - 2$ (b) $X^2 - 1$
    (c) $X^3 - 2$ (d) $(X^3 - 2)(X^2 - 2)$
    (e) $X^2 + X + 1$ (f) $X^6 + X^3 + 1$
    (g) $X^5 - 7$

12. Let K be a finite field with p^n elements. Show that every element of K has a unique p-th root in K.

13. If the roots of a monic polynomial $f(X) \in k[X]$ in some splitting field are distinct, and form a field, then char $k = p$ and $f(X) = X^{p^n} - X$ for some $n \geqq 1$.

14. Let char $K = p$. Let L be a finite extension of K, and suppose $[L:K]$ prime to p. Show that L is separable over K.

15. Suppose char $K = p$. Let $a \in K$. If a has no p-th root in K, show that $X^{p^n} - a$ is irreducible in $K[X]$ for all positive integers n.

16. Let char $K = p$. Let α be algebraic over K. Show that α is separable if and only if $K(\alpha) = K(\alpha^{p^n})$ for all positive integers n.

17. Prove that the following two properties are equivalent:
    (a) Every algebraic extension of K is separable.
    (b) Either char $K = 0$, or char $K = p$ and every element of K has a p-th root in K.

18. Show that every element of a finite field can be written as a sum of two squares in that field.

19. Let E be an algebraic extension of F. Show that every subring of E which contains F is actually a field. Is this necessarily true if E is not algebraic over F? Prove or give a counterexample.

20. (a) Let $E = F(x)$ where x is transcendental over F. Let $K \neq F$ be a subfield of E which contains F. Show that x is algebraic over K.
    (b) Let $E = F(x)$. Let $y = f(x)/g(x)$ be a rational function, with relatively prime polynomials f, $g \in F[x]$. Let $n = \max(\deg f, \deg g)$. Suppose $n \geqq 1$. Prove that

$$[F(x):F(y)] = n.$$

21. Let $\mathbf{Z}^+$ be the set of positive integers, and A an additive abelian group. Let $f: \mathbf{Z}^+ \to A$ and $g: \mathbf{Z}^+ \to A$ be maps. Suppose that for all n,

$$f(n) = \sum_{d|n} g(d).$$

    Let μ be the Möbius function (cf. Exercise 12 of Chapter II). Prove that

$$g(n) = \sum_{d|n} \mu(n/d) f(d).$$

22. Let k be a finite field with q elements. Let $f(X) \in k[X]$ be irreducible. Show that $f(X)$ divides $X^{q^n} - X$ if and only if deg f divides n. Show the multiplication formula

$$X^{q^n} - X = \prod_{d|n} \prod_{f_d \text{ irr}} f_d(X),$$

    where the inner product is over all irreducible polynomials of degree d with leading coefficient 1. Counting degrees, show that

$$q^n = \sum_{d|n} d\psi(d),$$

    where $\psi(d)$ is the number of irreducible polynomials of degree d. Invert by

Exercise 21 and find that

$$n\psi(n) = \sum_{d|n} \mu(d) q^{n/d}.$$

23. (a) Let k be a finite field with q elements. Define the **zeta function**

$$Z(t) = (1-t)^{-1} \prod_{p} (1 - t^{\deg p})^{-1},$$

where p ranges over all irreducible polynomials $p = p(X)$ in $k[X]$ with leading coefficient 1. Prove that $Z(t)$ is a rational function and determine this rational function.

(b) Let $\pi_q(n)$ be the number of primes p as in (a) of degree $\leqq n$. Prove that

$$\pi_q(m) \sim \frac{q}{q-1} \frac{q^m}{m} \qquad \text{for} \quad m \to \infty.$$

Remark. This is the analogue of the prime number theorem in number theory, but it is essentially trivial in the present case, because the Riemann hypothesis is trivially verified. Things get more interesting fast after this case. Consider an equation $y^2 = x^3 + ax + b$ over a finite field $\mathbf{F}_q$ of characteristic $\neq 2$, 3, and having q elements. Assume $-4a^3 - 27b^2 \neq 0$, in which case the curve defined by this equation is called an **elliptic curve**. Define N_n by

$N_n - 1 =$ number of points (x, y) satisfying the above equation with $x, y \in \mathbf{F}_{q^n}$ (the extension of $\mathbf{F}_q$ of degree n).

Define the **zeta function** $Z(t)$ to be the unique rational function such that $Z(0) = 1$ and

$$Z'/Z(t) = \sum N_n t^{n-1}.$$

A famous theorem of Hasse asserts that $Z(t)$ is a rational function of the form

$$Z(t) = \frac{(1 - \alpha t)(1 - \bar{\alpha} t)}{(1-t)(1-qt)},$$

where α is an imaginary quadratic number (not real, quadratic over $\mathbf{Q}$), $\bar{\alpha}$ is its complex conjugate, and $\alpha\bar{\alpha} = q$, so $|\alpha| = q^{1/2}$. See Hasse, "Abstrakte Bergrundung der komplexen Multiplikation und Riemannsche Vermutung in Funktionenkörpern," *Abh. Math. Sem. Univ. Hamburg* **10** (1934) pp. 325–348.

24. Let k be a field of characteristic p and let t, u be algebraically independent over k. Prove the following:
(a) $k(t, u)$ has degree p^2 over $k(t^p, u^p)$.
(b) There exist infinitely many extensions between $k(t, u)$ and $k(t^p, u^p)$.

25. Let E be a finite extension of k and let $p^r = [E : k]_i$. We assume that the characteristic is $p > 0$. Assume that there is no exponent p^s with $s < r$ such that $E^{p^s}k$ is separable over k (i.e., such that α^{p^s} is separable over k for each α in E). Show that E can be generated by one element over k. [*Hint*: Assume first that E is purely inseparable.]

26. Let k be a field, $f(X)$ an irreducible polynomial in $k[X]$, and let K be a finite normal extension of k. If g, h are monic irreducible factors of $f(X)$ in $K[X]$, show that there exists an automorphism σ of K over k such that $g = h^\sigma$. Give an example when this conclusion is not valid if K is not normal over k.

27. Let $x_1, \ldots, x_n$ be algebraically independent over a field k. Let y be algebraic over $k(x) = k(x_1, \ldots, x_n)$. Let $P(X_{n+1})$ be the irreducible polynomial of y over $k(x)$. Let $\varphi(x)$ be the least common multiple of the denominators of the coefficients of P. Then the coefficients of $\varphi(x)P$ are elements of $k[x]$. Show that the polynomial

$$f(X_1, \ldots, X_{n+1}) = \varphi(X_1, \ldots, X_n)P(X_{n+1})$$

is irreducible over k, as a polynomial in $n + 1$ variables.

Conversely, let $f(X_1, \ldots, X_{n+1})$ be an irreducible polynomial over k. Let $x_1, \ldots, x_n$ be algebraically independent over k. Show that

$$f(x_1, \ldots, x_n, X_{n+1})$$

is irreducible over $k(x_1, \ldots, x_n)$.

If f is a polynomial in n variables, and $(b) = (b_1, \ldots, b_n)$ is an n-tuple of elements such that $f(b) = 0$, then we say that (b) is a **zero** of f. We say that (b) is **non-trivial** if not all coordinates b_i are equal to 0.

28. Let $f(X_1, \ldots, X_n)$ be a homogeneous polynomial of degree 2 (resp. 3) over a field k. Show that if f has a non-trivial zero in an extension of odd degree (resp. degree 2) over k, then f has a non-trivial zero in k.

29. Let $f(X, Y)$ be an irreducible polynomial in two variables over a field k. Let t be transcendental over k, and assume that there exist integers m, $n \neq 0$ and elements $a, b \in k$, $ab \neq 0$, such that $f(at^n, bt^m) = 0$. Show that after inverting possibly X or Y, and up to a constant factor, f is of type

$$X^m Y^n - c$$

with some $c \in k$.

The answer to the following exercise is not known.

30. **(Artin conjecture).** Let f be a homogeneous polynomial of degree d in n variables, with rational coefficients. If $n > d$, show that there exists a root of unity ζ, and elements

$$x_1, \ldots, x_n \in \mathbf{Q}[\zeta]$$

not all 0 such that $f(x_1, \ldots, x_n) = 0$.

31. **Difference equations.** Let $u_1, \ldots, u_d$ be elements of a field K. We want to solve for infinite vectors $(x_0, x_1, \ldots, x_n, \ldots)$ satisfying

(*) $$x_n = u_1 x_{n-1} + \cdots + u_d x_{n-d} \quad \text{for} \quad n \geqq d.$$

Define the **characteristic polynomial** of the system to be

$$X^d - (u_1 X^{d-1} + \cdots + u_d) = f(X).$$

Suppose α is a root of f.

(a) Show that $x_n = \alpha^n$ $(n \geqq 0)$ is a solution of (*).

(b) Show that the set of solutions of (*) is a vector space of dimension d.

(c) Assume that the characteristic polynomial has d distinct roots $\alpha_1, \ldots, \alpha_d$. Show that the solutions $(\alpha_1^n), \ldots, (\alpha_d^n)$ form a basis for the space of solutions.

(d) Let $x_n = b_1\alpha_1^n + \cdots + b_d\alpha_d^n$ for $n \geqq 0$, show how to solve for $b_1, \ldots, b_d$ in terms of $\alpha_1, \ldots, \alpha_d$ and $x_0, \ldots, x_{d-1}$. (Use the Vandermonde determinant.)

(e) Under the conditions of (d), let $F(T) = \sum x_n T^n$. Show that $F(T)$ represents a rational function, and give its partial fraction decomposition.

32. Let $d = 2$ for simplicity. Given $a_0, a_1, u, v, w, t \in K$, we want to find the solutions of the system

$$a_n = ua_{n-1} - vta_{n-2} - t^n w \qquad \text{for} \quad n \geqq 2.$$

Let α_1, α_2 be the roots of the characteristic polynomial, that is

$$1 - uX + vtX^2 = (1 - \alpha_1 X)(1 - \alpha_2 X).$$

Assume that α_1, α_2 are distinct, and also distinct from t. Let

$$F(X) = \sum_{n=0}^{\infty} a_n X^n.$$

(a) Show that there exist elements A, B, C of K such that

$$F(X) = \frac{A}{1 - \alpha_1 X} + \frac{B}{1 - \alpha_2 X} + \frac{C}{1 - tX}.$$

(b) Show that there is a unique solution to the difference equation given by

$$a_n = A\alpha_1^n + B\alpha_2^n + Ct^n \qquad \text{for} \quad n \geqq 0.$$

(To see an application of this formalism to modular forms, as in the work of Manin, Mazur, and Swinnerton-Dyer, cf. my *Introduction to Modular Forms*, Springer-Verlag, New York, 1976, Chapter XII, §2.)

33. Let R be a ring which we assume entire for simplicity. Let

$$g(T) = T^d - a_{d-1}T^{d-1} - \cdots - a_0$$

be a polynomial in $R[T]$, and consider the equation

$$T^d = a_0 + a_1 T + \cdots + a_{d-1}T^{d-1}.$$

Let x be a root of $g(T)$.

(a) For any integer $n \geqq d$ there is a relation

$$x^n = a_{0,n} + a_{1,n}x + \cdots + a_{d-1,n}x^{d-1}$$

with coefficients $a_{i,j}$ in $\mathbf{Z}[a_0, \ldots, a_{d-1}] \subset R$.

(b) Let $F(T) \in R[T]$ be a polynomial. Then

$$F(x) = a_0(F) + a_1(F)x + \cdots + a_{d-1}(F)x^{d-1}$$

where the coefficients $a_i(F)$ lie in R and depend linearly on F.

(c) Let the Vandermonde determinant be

$$V(x_1, \ldots, x_d) = \begin{vmatrix} 1 & x_1 & \cdots & x_1^{d-1} \\ 1 & x_2 & \cdots & x_2^{d-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_d & \cdots & x_d^{d-1} \end{vmatrix} = \prod_{i<j} (x_j - x_i).$$

Suppose that the equation $g(T) = 0$ has d roots and that there is a factorization

$$g(T) = \prod_{i=1}^{d} (T - x_i).$$

Substituting x_i for x with $i = 1, \ldots, d$ and using Cramer's rule on the resulting system of linear equations, yields

$$\Delta a_j(F) = \Delta_j(F)$$

where Δ is the Vandermonde determinant, and $\Delta_j(F)$ is obtained by replacing the j-th column by ${}^t(F(x_1), \ldots, F(x_d))$, so

$$\Delta_j(F) = \begin{vmatrix} 1 & x_1 & \cdots & F(x_1) & \cdots & x_1^{d-1} \\ 1 & x_2 & \cdots & F(x_2) & \cdots & x_2^{d-1} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & x_d & \cdots & F(x_d) & \cdots & x_d^{d-1} \end{vmatrix}$$

If $\Delta \neq 0$ then we can write

$$a_j(F) = \Delta_j(F)/\Delta.$$

Remark. If $F(T)$ is a power series in $R[[T]]$ and if R is a complete local ring, with $x_1, \ldots, x_d$ in the maximal ideal, and $x = x_i$ for some i, then we can evaluate $F(x)$ because the series converges. The above formula for the coefficients $a_j(F)$ remains valid.

34. Let $x_1, \ldots, x_d$ be independent variables, and let A be the ring

$$\mathbf{Q}[[x_1, \ldots, x_d]][T]/\prod_{i=1}^{d} (T - x_i).$$

Substituting some x_i for T induces a natural homomorphism φ_i of A onto

$$\mathbf{Q}[[z_1, \ldots, x_d]] = R,$$

and the map $z \mapsto (\varphi_1(z), \ldots, \varphi_d(z))$ gives an embedding of A into the product of R with itself d times.

Let k be an integer, and consider the formal power series

$$F(T) = e^{kT} \prod_{i=1}^{d} \frac{(T - x_i)e^{T - x_i}}{e^{T - x_i} - 1} = e^{kT} \prod_{i=1}^{d} h(T - x_i)$$

where $h(t) = te^t/(e^t - 1)$. It is a formal power series in T, $T - x_1$, $\ldots$, $T - x_d$. Under substitution of some x_j for T it becomes a power series in x_j and $x_j - x_i$, and thus converges in $\mathbf{Q}[[x_1, \ldots, x_d]]$.

(a) Verify that

$$F(T) \equiv a_0(F) + \cdots + a_{d-1}(F)T^{d-1} \bmod \prod_{i=1}^{d} (T - x_i)$$

where $a_0(F), \ldots, a_{d-1}(F) \in \mathbf{Q}[[x_1, \ldots, x_d]]$, and that the formula given in the preceding exercise for these coefficients in terms of Vandermonde determinants is valid.

(b) Show that $a_{d-1}(F) = 0$ if $-(d-1) \leqq k < 0$ and $a_{d-1}(F) = 1$ if $k = 0$.

Remark. The assertion in (a) is a simple limit. The assertion in (b) is a fact which has been used in the proof of the Hirzebruch–Grothendieck–Riemann–Roch theorem and as far as I know there was no simple known proof until Roger Howe pointed out that it could be done by the formula of the preceding exercise as follows. We have

$$V(x_1, \ldots, x_n)a_{d-1}(F) = \begin{vmatrix} 1 & x_1 & \cdots & x_1^{d-2} & F(x_1) \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & x_d & \cdots & x_d^{d-2} & F(x_d) \end{vmatrix}.$$

Furthermore,

$$F(x_j) = e^{kx_j} \prod_{n \neq j} \frac{(x_j - x_n)e^{x_j - x_n}}{e^{x_j - x_n} - 1}.$$

We use the inductive relation of Vandermonde determinants

$$V(x_1, \ldots, x_d) = V(x_1, \ldots, \hat{x}_j, \ldots, x_d)(-1)^{d-j} \prod_{n \neq j} (x_j - x_n).$$

We expand the determinant for $a_{d-1}(F)$ according to the last column to get

$$a_{d-1}(F) = \sum_{j=1}^{d} e^{(k+d-1)x_j} \prod_{n \neq j} \frac{1}{e^{x_j} - e^{x_n}}.$$

Using the inductive relation backward, and replacing x_i by e^{x_i} which we denote by y_i for typographical reasons, we get

$$V(y_1, \ldots, y_d)a_{d-1}(F) = \begin{vmatrix} 1 & y_1 & \cdots & y_1^{d-2} & y_1^{k+d-1} \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & y_d & \cdots & y_d^{d-2} & y_d^{k+d-1} \end{vmatrix}$$

If $k \neq 0$ then two columns on the right are the same, so the determinant is 0. If $k = 0$ then we get the Vandermonde determinant on the right, so $a_{d-1}(F) = 1$. This proves the desired value.

CHAPTER VI

Galois Theory

This chapter contains the core of Galois theory. We study the group of automorphisms of a finite (and sometimes infinite) Galois extension at length, and give examples, such as cyclotomic extensions, abelian extensions, and even non-abelian ones, leading into the study of matrix representations of the Galois group and their classifications. We shall mention a number of fundamental unsolved problems, the most notable of which is whether given a finite group G, there exists a Galois extension of $\mathbf{Q}$ having this group as Galois group. Three surveys give recent points of view on those questions and sizeable bibliographies:

B. MATZAT, *Konstruktive Galoistheorie*, Springer Lecture Notes **1284**, 1987

B. MATZAT, Über das Umkehrproblem der Galoisschen Theorie, *Jahrsbericht Deutsch. Mat.-Verein.* **90** (1988), pp. 155–183

J. P. SERRE, *Topics in Galois theory*, course at Harvard, 1989, Jones and Bartlett, Boston 1992

More specific references will be given in the text at the appropriate moment concerning this problem and the problem of determining Galois groups over specific fields, especially the rational numbers.

§1. GALOIS EXTENSIONS

Let K be a field and let G be a group of automorphisms of K. We denote by K^G the subset of K consisting of all elements $x \in K$ such that $x^\sigma = x$ for all $\sigma \in G$. It is also called the **fixed field** of G. It is a field because if $x, y \in K^G$ then

$$(x + y)^\sigma = x^\sigma + y^\sigma = x + y$$

S. Lang, *Algebra*
© Springer Science+Business Media LLC 2002

for all $\sigma \in G$, and similarly, one verifies that K is closed under multiplication, subtraction, and multiplicative inverse. Furthermore, K^G contains 0 and 1, hence contains the prime field.

An algebraic extension K of a field k is called **Galois** if it is normal and separable. We consider K as embedded in an algebraic closure. The group of automorphisms of K over k is called the **Galois group** of K over k, and is denoted by $G(K/k)$, $G_{K/k}$, $\mathrm{Gal}(K/k)$, or simply G. It coincides with the set of embeddings of K in K^a over k.

For the convenience of the reader, we shall now state the main result of the Galois theory for finite Galois extensions.

Theorem 1.1. *Let K be a finite Galois extension of k, with Galois group G. There is a bijection between the set of subfields E of K containing k, and the set of subgroups H of G, given by $E = K^H$. The field E is Galois over k if and only if H is normal in G, and if that is the case, then the map $\sigma \mapsto \sigma | E$ induces an isomorphism of G/H onto the Galois group of E over k.*

We shall give the proofs step by step, and as far as possible, we give them for infinite extensions.

Theorem 1.2. *Let K be a Galois extension of k. Let G be its Galois group. Then $k = K^G$. If F is an intermediate field, $k \subset F \subset K$, then K is Galois over F. The map*

$$F \mapsto G(K/F)$$

from the set of intermediate fields into the set of subgroups of G is injective.

Proof. Let $\alpha \in K^G$. Let σ be any embedding of $k(\alpha)$ in K^a, inducing the identity on k. Extend σ to an embedding of K into K^a, and call this extension σ also. Then σ is an automorphism of K over k, hence is an element of G. By assumption, σ leaves α fixed. Therefore

$$[k(\alpha) : k]_s = 1.$$

Since α is separable over k, we have $k(\alpha) = k$ and α is an element of k. This proves our first assertion.

Let F be an intermediate field. Then K is normal and separable over F by Theorem 3.4 and Theorem 4.5 of Chapter V. Hence K is Galois over F. If $H = G(K/F)$ then by what we proved above we conclude that $F = K^H$. If F, F' are intermediate fields, and $H = G(K/F)$, $H' = G(K/F')$, then

$$F = K^H \quad \text{and} \quad F' = K^{H'}.$$

If $H = H'$ we conclude that $F = F'$, whence our map

$$F \mapsto G(K/F)$$

is injective, thereby proving our theorem.

We shall sometimes call the group $G(K/F)$ of an intermediate field the group **associated** with F. We say that a subgroup H of G **belongs** to an intermediate field F if $H = G(K/F)$.

Corollary 1.3. *Let K/k be Galois with group G. Let F, F' be two intermediate fields, and let H, H' be the subgroups of G belonging to F, F' respectively. Then $H \cap H'$ belongs to FF'.*

Proof. Every element of $H \cap H'$ leaves FF' fixed, and every element of G which leaves FF' fixed also leaves F and F' fixed and hence lies in $H \cap H'$. This proves our assertion.

Corollary 1.4. *Let the notation be as in Corollary* 1.3. *The fixed field of the smallest subgroup of G containing H, H' is $F \cap F'$.*

Proof. Obvious.

Corollary 1.5. *Let the notation be as in Corollary* 1.3. *Then $F \subset F'$ if and only if $H' \subset H$.*

Proof. If $F \subset F'$ and $\sigma \in H'$ leaves F' fixed then σ leaves F fixed, so σ lies in H. Conversely, if $H' \subset H$ then the fixed field of H is contained in the fixed field of H', so $F \subset F'$.

Corollary 1.6. *Let E be a finite separable extension of a field k. Let K be the smallest normal extension of k containing E. Then K is finite Galois over k. There is only a finite number of intermediate fields F such that $k \subset F \subset E$.*

Proof. We know that K is normal and separable, and K is finite over k since we saw that it is the finite compositum of the finite number of conjugates of E. The Galois group of K/k has only a finite number of subgroups. Hence there is only a finite number of subfields of K containing k, whence *a fortiori* a finite number of subfields of E containing k.

Of course, the last assertion of Corollary 1.6 has been proved in the preceding chapter, but we get another proof here from another point of view.

Lemma 1.7. *Let E be an algebraic separable extension of k. Assume that there is an integer $n \geqq 1$ such that every element α of E is of degree $\leqq n$ over k. Then E is finite over k and $[E : k] \leqq n$.*

Proof. Let α be an element of E such that the degree $[k(\alpha) : k]$ is maximal, say $m \leqq n$. We contend that $k(\alpha) = E$. If this is not true, then there exists an element $\beta \in E$ such that $\beta \notin k(\alpha)$, and by the primitive element theorem, there exists an element $\gamma \in k(\alpha, \beta)$ such that $k(\alpha, \beta) = k(\gamma)$. But from the tower

$$k \subset k(\alpha) \subset k(\alpha, \beta)$$

we see that $[k(\alpha, \beta) : k] > m$ whence γ has degree $> m$ over k, contradiction.

Theorem 1.8. (Artin). *Let K be a field and let G be a finite group of automorphisms of K, of order n. Let $k = K^G$ be the fixed field. Then K is a finite Galois extension of k, and its Galois group is G. We have $[K : k] = n$.*

Proof. Let $\alpha \in K$ and let $\sigma_1, \ldots, \sigma_r$ be a maximal set of elements of G such that $\sigma_1\alpha, \ldots, \sigma_r\alpha$ are distinct. If $\tau \in G$ then $(\tau\sigma_1\alpha, \ldots, \tau\sigma_r\alpha)$ differs from $(\sigma_1\alpha, \ldots, \sigma_r\alpha)$ by a permutation, because τ is injective, and every $\tau\sigma_i\alpha$ is among the set $\{\sigma_1\alpha, \ldots, \sigma_r\alpha\}$; otherwise this set is not maximal. Hence α is a root of the polynomial

$$f(X) = \prod_{i=1}^{r} (X - \sigma_i\alpha),$$

and for any $\tau \in G$, $f^\tau = f$. Hence the coefficients of f lie in $K^G = k$. Furthermore, f is separable. Hence every element α of K is a root of a separable polynomial of degree $\leqq n$ with coefficients in k. Furthermore, this polynomial splits in linear factors in K. Hence K is separable over k, is normal over k, hence Galois over k. By Lemma 1.7, we have $[K : k] \leqq n$. The Galois group of K over k has order $\leqq [K:k]$ (by Theorem 4.1 of Chapter V), and hence G must be the full Galois group. This proves all our assertions.

Corollary 1.9. *Let K be a finite Galois extension of k and let G be its Galois group. Then every subgroup of G belongs to some subfield F such that $k \subset F \subset K$.*

Proof. Let H be a subgroup of G and let $F = K^H$. By Artin's theorem we know that K is Galois over F with group H.

Remark. When K is an infinite Galois extension of k, then the preceding corollary is not true any more. This shows that some counting argument must be used in the proof of the finite case. In the present treatment, we have used an old-fashioned argument. The reader can look up Artin's own proof in his book *Galois Theory*. In the infinite case, one defines the Krull topology on the Galois group G (cf. exercises 43–45), and G becomes a compact totally disconnected group. The subgroups which belong to the intermediate fields are the *closed* subgroups. The reader may disregard the infinite case entirely throughout our discussions without impairing understanding. The proofs in the infinite case are usually identical with those in the finite case.

The notions of a Galois extension and a Galois group are defined completely algebraically. Hence they behave formally under isomorphisms the way one expects from objects in any category. We describe this behavior more explicitly in the present case.

Let K be a Galois extension of k. Let

$$\lambda : K \to \lambda K$$

be an isomorphism. Then λK is a Galois extension of λk.

$$\begin{array}{ccc} K & \xrightarrow{\lambda} & \lambda K \\ | & & | \\ k & \xrightarrow[\lambda]{} & \lambda k \end{array}$$

Let G be the Galois group of K over k. Then the map

$$\sigma \mapsto \lambda \circ \sigma \circ \lambda^{-1}$$

gives a homomorphism of G into the Galois group of λK over λk, whose inverse is given by

$$\lambda^{-1} \circ \tau \circ \lambda \longleftarrow\!\shortmid \tau.$$

Hence $G(\lambda K/\lambda k)$ is isomorphic to $G(K/k)$ under the above map. We may write

$$G(\lambda K/\lambda k)^{\lambda} = G(K/k)$$

or

$$G(\lambda K/\lambda k) = \lambda G(K/k)\lambda^{-1},$$

where the exponent λ is "conjugation,"

$$\sigma^{\lambda} = \lambda^{-1} \circ \sigma \circ \lambda.$$

There is no avoiding the contravariance if we wish to preserve the rule

$$(\sigma^{\lambda})^{\omega} = \sigma^{\lambda\omega}$$

when we compose mappings λ and ω.

In particular, let F be an intermediate field, $k \subset F \subset K$, and let $\lambda : F \to \lambda F$ be an embedding of F in K, which we assume is extended to an automorphism of K. Then $\lambda K = K$. Hence

$$G(K/\lambda F)^{\lambda} = G(K/F)$$

and

$$G(K/\lambda F) = \lambda G(K/F)\lambda^{-1}.$$

Theorem 1.10. *Let K be a Galois extension of k with group G. Let F be a subfield, $k \subset F \subset K$, and let $H = G(K/F)$. Then F is normal over k if and only if H is normal in G. If F is normal over k, then the restriction map $\sigma \mapsto \sigma | F$*

is a homomorphism of G onto the Galois group of F over k, whose kernel is H. We thus have $G(F/k) \approx G/H$.

Proof. Assume F is normal over k, and let G' be its Galois group. The restriction map $\sigma \to \sigma|F$ maps G into G', and by definition, its kernel is H. Hence H is normal in G. Furthermore, any element $\tau \in G'$ extends to an embedding of K in K^a, which must be an automorphism of K, so the restriction map is surjective. This proves the last statement. Finally, assume that F is not normal over k. Then there exists an embedding λ of F in K over k which is not an automorphism, i.e. $\lambda F \neq F$. Extend λ to an automorphism of K over k. The Galois groups $G(K/\lambda F)$ and $G(K/F)$ are conjugate, and they belong to distinct subfields, hence cannot be equal. Hence H is not normal in G.

A Galois extension K/k is said to be **abelian** (resp. **cyclic**) if its Galois group G is abelian (resp. cyclic).

Corollary 1.11. *Let K/k be abelian (resp. cyclic). If F is an intermediate field, $k \subset F \subset K$, then F is Galois over k and abelian (resp. cyclic).*

Proof. This follows at once from the fact that a subgroup of an abelian group is normal, and a factor group of an abelian (resp. cyclic) group is abelian (resp. cyclic).

Theorem 1.12. *Let K be a Galois extension of k, let F be an arbitrary extension and assume that K, F are subfields of some other field. Then KF is Galois over F, and K is Galois over $K \cap F$. Let H be the Galois group of KF over F, and G the Galois group of K over k. If $\sigma \in H$ then the restriction of σ to K is in G, and the map*

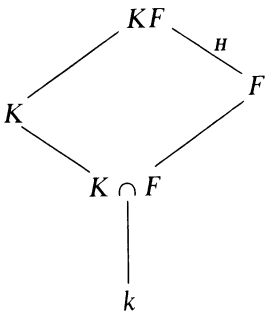
$$\sigma \mapsto \sigma|K$$

gives an isomorphism of H on the Galois group of K over $K \cap F$.

Proof. Let $\sigma \in H$. The restriction of σ to K is an embedding of K over k, whence an element of G since K is normal over k. The map $\sigma \mapsto \sigma|K$ is clearly a homomorphism. If $\sigma|K$ is the identity, then σ must be the identity of KF (since every element of KF can be expressed as a combination of sums, products, and quotients of elements in K and F). Hence our homomorphism $\sigma \mapsto \sigma|K$ is injective. Let H' be its image. Then H' leaves $K \cap F$ fixed, and conversely, if an element $\alpha \in K$ is fixed under H', we see that α is also fixed under H, whence $\alpha \in F$ and $\alpha \in K \cap F$. Therefore $K \cap F$ is the fixed field. If K is finite over k, or even KF finite over F, then by Theorem 1.8, we know that H' is the Galois group of K over $K \cap F$, and the theorem is proved in that case.

(In the infinite case, one must add the remark that for the Krull topology, our map $\sigma \mapsto \sigma|K$ is continuous, whence its image is closed since H is compact. See Theorem 14.1; Chapter I, Theorem 10.1; and Exercise 43.)

The diagram illustrating Theorem 1.12 is as follows:



It is suggestive to think of the opposite sides of a parallelogram as being equal.

Corollary 1.13. *Let K be a finite Galois extension of k. Let F be an arbitrary extension of k. Then $[KF : F]$ divides $[K : k]$.*

Proof. Notation being as above, we know that the order of H divides the order of G, so our assertion follows.

Warning. The assertion of the corollary is not usually valid if K is not Galois over k. For instance, let $\alpha = \sqrt[3]{2}$ be the real cube root of 2, let ζ be a cube root of 1, $\zeta \neq 1$, say

$$\zeta = \frac{-1 + \sqrt{-3}}{2},$$

and let $\beta = \zeta\alpha$. Let $E = \mathbf{Q}(\beta)$. Since β is complex and α real, we have

$$\mathbf{Q}(\beta) \neq \mathbf{Q}(\alpha).$$

Let $F = \mathbf{Q}(\alpha)$. Then $E \cap F$ is a subfield of E whose degree over $\mathbf{Q}$ divides 3. Hence this degree is 3 or 1, and must be 1 since $E \neq F$. But

$$EF = \mathbf{Q}(\alpha, \beta) = \mathbf{Q}(\alpha, \zeta) = \mathbf{Q}(\alpha, \sqrt{-3}).$$

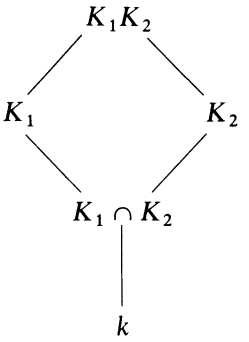
Hence EF has degree 2 over F.

Theorem 1.14. *Let K_1 and K_2 be Galois extensions of a field k, with Galois groups G_1 and G_2 respectively. Assume K_1, K_2 are subfields of some field. Then K_1K_2 is Galois over k. Let G be its Galois group. Map $G \to G_1 \times G_2$ by restriction, namely*

$$\sigma \mapsto (\sigma | K_1, \sigma | K_2).$$

This map is injective. If $K_1 \cap K_2 = k$ then the map is an isomorphism.

Proof. Normality and separability are preserved in taking the compositum of two fields, so K_1K_2 is Galois over k. Our map is obviously a homomorphism of G into $G_1 \times G_2$. If an element $\sigma \in G$ induces the identity on K_1 and K_2 then it induces the identity on their compositum, so our map is injective. Assume that $K_1 \cap K_2 = k$. According to Theorem 1.12, given an element $\sigma_1 \in G_1$ there exists an element σ of the Galois group of K_1K_2 over K_2 which induces σ_1 on K_1. This σ is *a fortiori* in G, and induces the identity on K_2. Hence $G_1 \times \{e_2\}$ is contained in the image of our homomorphism (where e_2 is the unit element of G_2). Similarly, $\{e_1\} \times G_2$ is contained in this image. Hence their product is contained in the image, and their product is precisely $G_1 \times G_2$. This proves Theorem 1.14.



Corollary 1.15. *Let $K_1, \ldots, K_n$ be Galois extensions of k with Galois groups $G_1, \ldots, G_n$. Assume that $K_{i+1} \cap (K_1 \cdots K_i) = k$ for each $i = 1, \ldots, n-1$. Then the Galois group of $K_1 \cdots K_n$ is isomorphic to the product $G_1 \times \cdots \times G_n$ in the natural way.*

Proof. Induction.

Corollary 1.16. *Let K be a finite Galois extension of k with group G, and assume that G can be written as a direct product $G = G_1 \times \cdots \times G_n$. Let K_i be the fixed field of*

$$G_1 \times \cdots \times \{1\} \times \cdots \times G_n$$

where the group with 1 element occurs in the i-th place. Then K_i is Galois over k, and $K_{i+1} \cap (K_1 \cdots K_i) = k$. Furthermore $K = K_1 \cdots K_n$.

Proof. By Corollary 1.3, the compositum of all K_i belongs to the intersection of their corresponding groups, which is clearly the identity. Hence the compositum is equal to K. Each factor of G is normal in G, so K_i is Galois over k. By Corollary 1.4, the intersection of normal extensions belongs to the product of their Galois groups, and it is then clear that $K_{i+1} \cap (K_1 \cdots K_i) = k$.

Theorem 1.17. *Assume all fields contained in some common field.*

(i) *If K, L are abelian over k, so is the composite KL.*

(ii) *If K is abelian over k and E is any extension of k, then KE is abelian over E.*

(iii) *If K is abelian over k and $K \supset E \supset k$ where E is an intermediate field, then E is abelian over k and K is abelian over E.*

Proof. Immediate from Theorems 1.12 and 1.14.

If k is a field, the compositum of all abelian extensions of k in a given algebraic closure k^a is called the **maximum abelian extension** of k, and is denoted by k^{ab}.

Remark on notation. We have used systematically the notation:

k^a = algebraic closure of k;

k^s = separable closure of k;

k^{ab} = abelian closure of k = maximal abelian extension.

We have replaced other people's notation $\bar{k}$ (and mine as well in the first edition) with k^a in order to make the notation functorial with respect to the ideas.

§2. EXAMPLES AND APPLICATIONS

Let k be a field and $f(X)$ a separable polynomial of degree $\geqq 1$ in $k[X]$. Let

$$f(X) = (X - \alpha_1) \cdots (X - \alpha_n)$$

be its factorization in a splitting field K over k. Let G be the Galois group of K over k. We call G the **Galois group** of f over k. Then the elements of G permute the roots of f. Thus we have an injective homomorphism of G into the symmetric group S_n on n elements. Not every permutation need be given by an element of G. We shall discuss examples below.

Example 1. Quadratic extensions. Let k be a field and $a \in k$. If a is not a square in k, then the polynomial $X^2 - a$ has no root in k and is therefore irreducible. Assume char $k \neq 2$. Then the polynomial is separable (because $2 \neq 0$), and if α is a root, then $k(\alpha)$ is the splitting field, is Galois, and its Galois group is cyclic of order 2.

Conversely, given an extension K of k of degree 2, there exists $a \in k$ such that $K = k(\alpha)$ and $\alpha^2 = a$. This comes from completing the square and the quadratic formula as in elementary school. The formula is valid as long as the characteristic of k is $\neq 2$.

Example 2. Cubic extensions. Let k be a field of characteristic $\neq 2$ or 3. Let

$$f(X) = X^3 + aX + b.$$

Any polynomial of degree 3 can be brought into this form by completing the cube. Assume that f has no root in k. Then f is irreducible because any factorization must have a factor of degree 1. Let α be a root of $f(X)$. Then

$$[k(\alpha): k] = 3.$$

Let K be the splitting field. Since char $k \neq 2, 3$, f is separable. Let G be the Galois group. Then G has order 3 or 6 since G is a subgroup of the symmetric group S_3. In the second case, $k(\alpha)$ is not normal over k.

There is an easy way to test whether the Galois group is the full symmetric group. We consider the discriminant. If $\alpha_1, \alpha_2, \alpha_3$ are the distinct roots of $f(X)$, we let

$$\delta = (\alpha_1 - \alpha_2)(\alpha_2 - \alpha_3)(\alpha_1 - \alpha_3) \quad \text{and} \quad \Delta = \delta^2.$$

If G is the Galois group and $\sigma \in G$ then $\sigma(\delta) = \pm\delta$. Hence σ leaves Δ fixed. Thus Δ is in the ground field k, and in Chapter IV, §6, we have seen that

$$\Delta = -4a^3 - 27b^2.$$

The set of σ in G which leave δ fixed is precisely the set of even permutations. Thus G is the symmetric group if and only if Δ is not a square in k. We may summarize the above remarks as follows.

Let $f(X)$ be a cubic polynomial in $k[X]$, and assume char $k \neq 2, 3$. *Then:*

(a) *f is irreducible over k if and only if f has no root in k.*

(b) *Assume f irreducible. Then the Galois group of f is S_3 if and only if the discriminant of f is not a square in k. If the discriminant is a square, then the Galois group is cyclic of order* 3, *equal to the alternating group A_3 as a permutation of the roots of f.*

For instance, consider

$$f(X) = X^3 - X + 1$$

over the rational numbers. Any rational root must be 1 or -1, and so $f(X)$ is irreducible over $\mathbf{Q}$. The discriminant is -23, and is not a square. Hence the Galois group is the symmetric group. The splitting field contains a subfield of degree 2, namely $k(\delta) = k(\sqrt{\Delta})$.

On the other hand, let $f(X) = X^3 - 3X + 1$. Then f has no root in $\mathbf{Z}$, whence no root in $\mathbf{Q}$, so f is irreducible. The discriminant is 81, which is a square, so the Galois group is cyclic of order 3.

Example 3. We consider the polynomial $f(X) = X^4 - 2$ over the rationals $\mathbf{Q}$. It is irreducible by Eisenstein's criterion. Let α be a real root.

Let $i = \sqrt{-1}$. Then $\pm\alpha$ and $\pm i\alpha$ are the four roots of $f(X)$, and

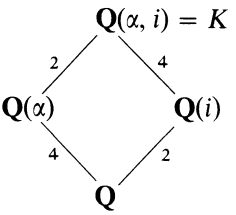
$$[\mathbf{Q}(\alpha) : \mathbf{Q}] = 4.$$

Hence the splitting field of $f(X)$ is

$$K = \mathbf{Q}(\alpha, i).$$

The field $\mathbf{Q}(\alpha) \cap \mathbf{Q}(i)$ has degree 1 or 2 over $\mathbf{Q}$. The degree cannot be 2 otherwise $i \in \mathbf{Q}(\alpha)$, which is impossible since α is real. Hence the degree is 1. Hence i has degree 2 over $\mathbf{Q}(\alpha)$ and therefore $[K : \mathbf{Q}] = 8$. The Galois group of $f(X)$ has order 8.

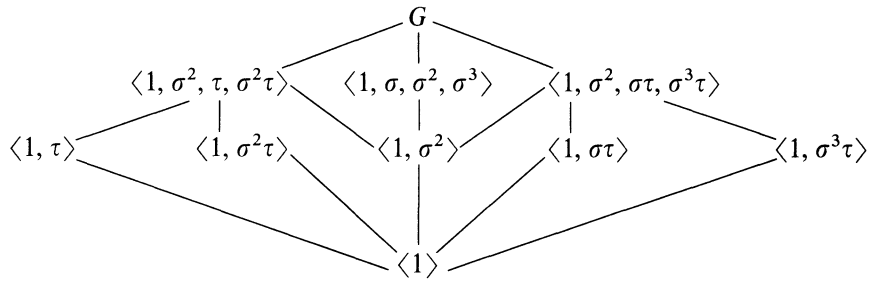
There exists an automorphism τ of K leaving $\mathbf{Q}(\alpha)$ fixed, sending i to $-i$, because K is Galois over $\mathbf{Q}(\alpha)$, of degree 2. Then $\tau^2 = \mathrm{id}$.



By the multiplicativity of degrees in towers, we see that the degrees are as indicated in the diagram. Thus $X^4 - 2$ is irreducible over $\mathbf{Q}(i)$. Also, K is normal over $\mathbf{Q}(i)$. There exists an automorphism σ of K over $\mathbf{Q}(i)$ mapping the root α of $X^4 - 2$ to the root $i\alpha$. Then one verifies at once that $1, \sigma, \sigma^2, \sigma^3$ are distinct and $\sigma^4 = \mathrm{id}$. Thus σ generates a cyclic group of order 4. We denote it by $\langle\sigma\rangle$. Since $\tau \notin \langle\sigma\rangle$ it follows that $G = \langle\sigma, \tau\rangle$ is generated by σ and τ because $\langle\sigma\rangle$ has index 2. Furthermore, one verifies directly that

$$\tau\sigma = \sigma^3\tau,$$

because this relation is true when applied to α and i which generate K over $\mathbf{Q}$. This gives us the structure of G. It is then easy to verify that the lattice of subgroups is as follows:



Example 4. Let k be a field and let $t_1, \ldots, t_n$ be algebraically independent over k. Let $K = k(t_1, \ldots, t_n)$. The symmetric group G on n letters operates on K by permuting $(t_1, \ldots, t_n)$ and its fixed field is the field of symmetric functions, by definition the field of those elements of K fixed under G. Let $s_1, \ldots, s_n$ be the elementary symmetric polynomials, and let

$$f(X) = \prod_{i=1}^{n} (X - t_i).$$

Up to a sign, the coefficients of f are $s_1, \ldots, s_n$. We let $F = K^G$. We contend that $F = k(s_1, \ldots, s_n)$. Indeed,

$$k(s_1, \ldots, s_n) \subset F.$$

On the other hand, K is the splitting field of $f(X)$, and its degree over F is $n!$. Its degree over $k(s_1, \ldots, s_n)$ is $\leqq n!$ and hence we have equality, $F = k(s_1, \ldots, s_n)$.

The polynomial $f(X)$ above is called the general polynomial of degree n. We have just constructed a Galois extension whose Galois group is the symmetric group.

Using the Hilbert irreducibility theorem, one can construct a Galois extension of $\mathbf{Q}$ whose Galois group is the symmetric group. (Cf. Chapter VII, end of §2, and [La 83], Chapter IX.) It is unknown whether given a finite group G, there exists a Galois extension of $\mathbf{Q}$ whose Galois group is G. By specializing parameters, Emmy Noether remarked that one could prove this if one knew that every field E such that

$$\mathbf{Q}(s_1, \ldots, s_n) \subset E \subset \mathbf{Q}(t_1, \ldots, t_n)$$

is isomorphic to a field generated by n algebraically independent elements. However, matters are not so simple, because Swan proved that the fixed field of a cyclic subgroup of the symmetric group is not necessarily generated by algebraically independent elements over k [Sw 69], [Sw 83].

Example 5. *We shall prove that the complex numbers are algebraically closed.* This will illustrate almost all the theorems we have proved previously.

We use the following properties of the real numbers $\mathbf{R}$: It is an ordered field, every positive element is a square, and every polynomial of odd degree in $\mathbf{R}[X]$ has a root in $\mathbf{R}$. We shall discuss ordered fields in general later, and our arguments apply to any ordered field having the above properties.

Let $i = \sqrt{-1}$ (in other words a root of $X^2 + 1$). Every element in $\mathbf{R}(i)$ has a square root. If $a + bi \in \mathbf{R}(i)$, $a, b \in \mathbf{R}$, then the square root is given by $c + di$, where

$$c^2 = \frac{a + \sqrt{a^2 + b^2}}{2} \quad \text{and} \quad d^2 = \frac{-a + \sqrt{a^2 + b^2}}{2}.$$

Each element on the right of our equalities is positive and hence has a square root in $\mathbf{R}$. It is then trivial to determine the sign of c and d so that $(c + di)^2 = a + bi$.

Since $\mathbf{R}$ has characteristic 0, every finite extension is separable. Every finite extension of $\mathbf{R}(i)$ is contained in an extension K which is finite and Galois over $\mathbf{R}$. We must show that $K = \mathbf{R}(i)$. Let G be the Galois group over $\mathbf{R}$ and let H be a 2-Sylow subgroup of G. Let F be its fixed field. Counting degrees and orders, we find that the degree of F over $\mathbf{R}$ is odd. By the primitive element theorem, there exists an element $\alpha \in F$ such that $F = \mathbf{R}(\alpha)$. Then α is the root of an irreducible polynomial in $\mathbf{R}[X]$ of odd degree. This can happen only if this degree is 1. Hence $G = H$ is a 2-group.

We now see that K is Galois over $\mathbf{R}(i)$. Let G_1 be its Galois group. Since G_1 is a p-group (with $p = 2$), if G_1 is not the trivial group, then G_1 has a subgroup G_2 of index 2. Let F be the fixed field of G_2. Then F is of degree 2 over $\mathbf{R}(i)$; it is a quadratic extension. But we saw that every element of $\mathbf{R}(i)$ has a square root, and hence that $\mathbf{R}(i)$ has no extensions of degree 2. It follows that G_1 is the trivial group and $K = \mathbf{R}(i)$, which is what we wanted.

(The basic ideas of the above proof were already in Gauss. The variation of the ideas which we have selected, making a particularly efficient use of the Sylow group, is due to Artin.)

Example 6. Let $f(X)$ be an irreducible polynomial over the field k, and assume that f is separable. Then the Galois group G of the splitting field is represented as a group of permutations of the n roots, where $n = \deg f$. Whenever one has a criterion for this group to be the full symmetric group S_n, then one can see if it applies to this representation of G. For example, it is an easy exercise (cf. Chapter I, Exercise 38) that for p prime, S_p is generated by $[123 \cdots p]$ and any transposition. We then have the following result.

Let $f(X)$ be an irreducible polynomial with rational coefficients and of degree p prime. If f has precisely two nonreal roots in the complex numbers, then the Galois group of f is S_p.

Proof. The order of G is divisible by p, and hence by Sylow's theorem, G contains an element of order p. Since G is a subgroup of S_p which has order $p!$, it follows that an element of order p can be represented by a p-cycle $[123 \cdots p]$ after a suitable ordering of the roots, because any smaller cycle has order less than p, so relatively prime to p. But the pair of complex conjugate roots shows that complex conjugation induces a transposition in G. Hence the group is all of S_p.

A specific case is easily given. Drawing the graph of

$$f(X) = X^5 - 4X + 2$$

shows that f has exactly three real roots, so exactly two complex conjugate roots. Furthermore f is irreducible over $\mathbf{Q}$ by Eisenstein's criterion, so we can apply the general statement proved above to conclude that the Galois group of f over $\mathbf{Q}$ is S_5. See also Exercise 17 of Chapter IV.

Example 7. The preceding example determines a Galois group by finding some subgroups passing to an extension field of the ground field. There are other possible extensions of $\mathbf{Q}$ rather than the reals, for instance p-adic fields which will be discussed later in this book. However, instead of passing to an extension field, it is possible to use reduction mod p. For our purposes here, we assume the following statement, which will be proved in Chapter VII, theorem 2.9.

Let $f(X) \in \mathbf{Z}[X]$ be a polynomial with integral coefficients, and leading coefficient 1. *Let p be a prime number. Let $\bar{f}(X) = f(X) \bmod p$ be the polynomial obtained by reducing the coefficients* mod p. *Assume that $\bar{f}$ has no multiple roots in an algebraic closure of $\mathbf{F}_p$. Then there exists a bijection*

$$(\alpha_1, \ldots, \alpha_n) \mapsto (\bar{\alpha}_1, \ldots, \bar{\alpha}_n)$$

of the roots of f onto those of $\bar{f}$, and an embedding of the Galois group of $\bar{f}$ as a subgroup of the Galois group of f, which gives an isomorphism of the action of those groups on the set of roots.

The embedding will be made precise in Chapter VII, but here we just want to use this result to compute Galois groups.

For instance, consider $X^5 - X - 1$ over $\mathbf{Z}$. Reducing mod 5 shows that this polynomial is irreducible. Reducing mod 2 gives the irreducible factors

$$(X^2 + X + 1)(X^3 + X^2 + 1) \pmod 2.$$

Hence the Galois group over the rationals contains a 5-cycle and a product of a 2-cycle and a 3-cycle. The third power of the product of the 2-cycle and 3-cycle is a 2-cycle, which is a transposition. Hence the Galois group contains a transposition and the cycle [12345], which generate S_5 (cf. the exercises of Chapter I on the symmetric group). Thus the Galois group of $X^5 - X - 1$ is S_5.

Example 8. The technique of reducing mod primes to get lots of elements in a Galois group was used by Schur to determine the Galois groups of classical polynomials [Schur 31]. For instance, Schur proves that the Galois group over $\mathbf{Q}$ of the following polynomials over $\mathbf{Q}$ is the symmetric group:

(a) $f(X) = \sum_{m=0}^{n} X^m/m!$ (in other words, the truncated exponential series), if n is not divisible by 4. If n is divisible by 4, he gets the alternating group.

(b) Let

$$H_m(X) = (-1)^m e^{X^2/2} \frac{d^m}{dX^m}(e^{-X^2/2})$$

be the m-th Hermite polynomial. Put

$$H_{2n}(X) = K_n^{(0)}(X^2) \quad \text{and} \quad H_{2n+1}(X) = XK_n^{(1)}(X^2).$$

Then the Galois group of $K_n^{(i)}(X)$ over $\mathbf{Q}$ is the symmetric group S_n for $i = 0$, 1, provided $n > 12$. The remaining cases were settled in [Schulz 37].

Example 9. This example is addressed to those who know something about Riemann surfaces and coverings. Let t be transcendental over the complex numbers $\mathbf{C}$, and let $k = \mathbf{C}(t)$. The values of t in $\mathbf{C}$, or ∞, correspond to the points of the Gauss sphere S, viewed as a Riemann surface. Let $P_1, \ldots, P_{n+1}$ be distinct points of S. The finite coverings of $S - \{P_1, \ldots, P_{n+1}\}$ are in bijection with certain finite extensions of $\mathbf{C}(t)$, those which are unramified outside $P_1, \ldots, P_{n+1}$. Let K be the union of all these extension fields corresponding to such coverings, and let $\pi_1^{(n)}$ be the fundamental group of

$$S - \{P_1, \ldots, P_{n+1}\}.$$

Then it is known that $\pi_1^{(n)}$ is a free group on n generators, and has an embedding in the Galois group of K over $\mathbf{C}(t)$, such that the finite subfields of K over $\mathbf{C}(t)$ are in bijection with the subgroups of $\pi_1^{(n)}$ which are of finite index. Given a finite group G generated by n elements $\sigma_1, \ldots, \sigma_n$ we can find a surjective homomorphism $\pi_1^{(n)} \to G$ mapping the generators of $\pi_1^{(n)}$ on $\sigma_1, \ldots, \sigma_n$. Let H be the kernel. Then H belongs to a subfield K^H of K which is normal over $\mathbf{C}(t)$ and whose Galois group is G. In the language of coverings, H belongs to a finite covering of

$$S - \{P_1, \ldots, P_{n+1}\}.$$

Over the field $\mathbf{C}(t)$ one can use analytic techniques to determine the Galois group. The Galois group is the completion of a free group, as proved by Douady [Dou 64]. For extensions to characteristic p, see [Pop 95]. A fundamental problem is to determine the Galois group over $\mathbf{Q}(t)$, which requires much deeper insight into the number theoretic nature of this field. Basic contributions were made by Belyi [Be 80], [Be 83], who also considered the field $\mathbf{Q}(\mu)(t)$, where $\mathbf{Q}(\mu)$ is the field obtained by adjoining all roots of unity to the rationals. Belyi proved that over this latter field, essentially all the classical finite groups occur as Galois groups. See also Conjecture 14.2 below.

For Galois groups over $\mathbf{Q}(t)$, see the survey [Se 88], which contains a bibliography. One method is called the rigidity method, first applied by Shih [Shi 74], which I summarize because it gives examples of various notions defined throughout this book. The problem is to descend extensions of $\mathbf{C}(t)$ with a given Galois group G to extensions of $\mathbf{Q}(t)$ with the same Galois group. If this extension is K over $\mathbf{Q}(t)$, one also wants the extension to be regular over $\mathbf{Q}$ (see the definition in Chapter VIII, §4). To give a sufficient condition, we need some definitions. Let G be a finite group with trivial center. Let C_1, C_2, C_3 be conjugacy classes. Let $P = P(C_1, C_2, C_3)$ be the set of elements

$$(g_1, g_2, g_3) \in C_1 \times C_2 \times C_3$$

such that $g_1 g_2 g_3 = 1$. Let P' be the subset of P consisting of all elements $(g_1, g_2, g_3) \in P$ such that G is generated by g_1, g_2, g_3. We say that the family (C_1, C_2, C_3) is **rigid** if G operates transitively on P', and P' is not empty.

We define a conjugacy class C of G to be **rational** if given $g \in C$ and a positive integer s relatively prime to the order of g, then $g^s \in C$. (Assuming that the reader knows the terminology of characters defined in Chapter XVIII, this condition of rationality is equivalent to the condition that every character χ of G has values in the rational numbers $\mathbf{Q}$.) One then has the following theorem, which is contained in the works of Shih, Fried, Belyi, Matzat and Thompson.

Rigidity theorem. *Let G be a finite group with trivial center, and let C_1, C_2, C_3 be conjugacy classes which are rational, and such that the family (C_1, C_2, C_3) is rigid. Then there exists a Galois extension of $\mathbf{Q}(t)$ with Galois group G (and such that the extension is regular over $\mathbf{Q}$).*

Bibliography

[Be 80] G. BELYI, Galois extensions of the maximal cyclotomic field, *Izv. Akad. Nauk SSR* **43** (1979) pp. 267–276 (= *Math. USSR Izv.* **14** (1980), pp. 247–256

[Be 83] G. BELYI, On extensions of the maximal cyclotomic field having a given classical Galois group, *J. reine angew. Math.* **341** (1983), pp. 147–156

[Dou 64] A. DOUADY, Determination d'un groupe de Galois, *C.R. Acad. Sci.* **258** (1964), pp. 5305–5308

[La 83] S. LANG, *Fundamentals of Diophantine Geometry*. Springer Verlag 1983

[Pop 95] F. POP, Etale Galois covers of affine smooth curves, *Invent. Math.* **120** (1995), pp. 555–578

[Se 88] J.-P.. SERRE, Groupes de Galois sur $\mathbf{Q}$, *Séminaire Bourbaki*, 1987–1988 *Astérisque* **161–162**, pp. 73–85

[Shi 74] R.-Y. SHIH, On the construction of Galois extensions of function fields and number fields, *Math. Ann.* **207** (1974), pp. 99–120

[Sw 69] R. SWAN, Invariant rational functions and a problem of Steenrod, *Invent. Math.* **7** (1969), pp. 148–158

[Sw 83] R. SWAN, Noether's problem in Galois theory, *Emmy Noether in Bryn Mawr*, J. D. Sally and B. Srinivasan, eds., Springer Verlag, 1983, pp. 40

§3. ROOTS OF UNITY

Let k be a field. By a **root of unity** (in k) we shall mean an element $\zeta \in k$ such that $\zeta^n = 1$ for some integer $n \geqq 1$. If the characteristic of k is p, then the equation

$$X^{p^m} = 1$$

has only one root, namely 1, and hence there is no p^m-th root of unity except 1.

Let n be an integer > 1 and not divisible by the characteristic. The polynomial

$$X^n - 1$$

is separable because its derivative is $nX^{n-1} \neq 0$, and the only root of the derivative is 0, so there is no common root. Hence in k^a the polynomial $X^n - 1$ has n distinct roots, which are roots of unity. They obviously form a group, and we know that every finite multiplicative group in a field is cyclic (Chapter IV, Theorem 1.9). Thus the group of n-th roots of unity is cyclic. A generator for this group is called a **primitive** n-th root of unity.

If $\boldsymbol{\mu}_n$ denotes the group of all n-th roots of unity in k^a and m, n are relatively prime integers, then

$$\boldsymbol{\mu}_{mn} \approx \boldsymbol{\mu}_m \times \boldsymbol{\mu}_n.$$

This follows because $\boldsymbol{\mu}_m$, $\boldsymbol{\mu}_n$ cannot have any element in common except 1, and because $\boldsymbol{\mu}_m\boldsymbol{\mu}_n$ consequently has mn elements, each of which is an mn-th root of unity. Hence $\boldsymbol{\mu}_m\boldsymbol{\mu}_n = \boldsymbol{\mu}_{mn}$, and the decomposition is that of a direct product.

As a matter of notation, to avoid double indices, especially in the prime power case, we write $\boldsymbol{\mu}[n]$ for $\boldsymbol{\mu}_n$. So if p is a prime, $\boldsymbol{\mu}[p^r]$ is the group of p^r-th roots of unity. Then $\boldsymbol{\mu}[p^\infty]$ denotes the union of all $\boldsymbol{\mu}[p^r]$ for all positive integers r. See the comments in §14.

Let k be any field. Let n be not divisible by the characteristic p. Let $\zeta = \zeta_n$ be a primitive n-th root of unity in k^a. Let σ be an embedding of $k(\zeta)$ in k^a over k. Then

$$(\sigma\zeta)^n = \sigma(\zeta^n) = 1$$

so that $\sigma\zeta$ is an n-th root of unity also. Hence $\sigma\zeta = \zeta^i$ for some integer $i = i(\sigma)$, uniquely determined mod n. It follows that σ maps $k(\zeta)$ into itself, and hence that $k(\zeta)$ is normal over k. If τ is another automorphism of $k(\zeta)$ over k then

$$\sigma\tau\zeta = \zeta^{i(\sigma)i(\tau)}.$$

Since σ and τ are automorphisms, it follows that $i(\sigma)$ and $i(\tau)$ are prime to n (otherwise, $\sigma\zeta$ would have a period smaller than n). In this way we get a homomorphism of the Galois group G of $k(\zeta)$ over k into the multiplicative group $(\mathbf{Z}/n\mathbf{Z})^*$ of integers prime to n, mod n. Our homomorphism is clearly injective since $i(\sigma)$ is uniquely determined by σ mod n, and the effect of σ on $k(\zeta)$ is determined by its effect on ζ. *We conclude that $k(\zeta)$ is abelian over k.*

We know that the order of $(\mathbf{Z}/n\mathbf{Z})^*$ is $\varphi(n)$. Hence the degree $[k(\zeta):k]$ divides $\varphi(n)$.

For a specific field k, the question arises whether the image of $G_{k(\zeta)/k}$ in $(\mathbf{Z}/n\mathbf{Z})^*$ is all of $(\mathbf{Z}/n\mathbf{Z})^*$. Looking at $k = \mathbf{R}$ or $\mathbf{C}$, one sees that this is not always the case. We now give an important example when it is the case.

Theorem 3.1. *Let ζ be a primitive n-th root of unity. Then*

$$[\mathbf{Q}(\zeta):\mathbf{Q}] = \varphi(n),$$

where φ is the Euler function. The map $\sigma \mapsto i(\sigma)$ gives an isomorphism

$$G_{\mathbf{Q}(\zeta)/\mathbf{Q}} \xrightarrow{\sim} (\mathbf{Z}/n\mathbf{Z})^*.$$

Proof. Let $f(X)$ be the irreducible polynomial of ζ over $\mathbf{Q}$. Then $f(X)$ divides $X^n - 1$, say $X^n - 1 = f(X)h(X)$, where both f, h have leading coefficient 1. By the Gauss lemma, it follows that f, h have integral coefficients. We shall now prove that if p is a prime number not dividing n, then ζ^p is also a root of f. Since ζ^p is also a primitive n-th root of unity, and since any primitive n-th root of unity can be obtained by raising ζ to a succession of prime powers, with primes not dividing n, this will imply that all the primitive n-th roots of unity are roots of f, which must therefore have degree $\geqq \varphi(n)$, and hence precisely $\varphi(n)$.

Suppose ζ^p is not a root of f. Then ζ^p is a root of h, and ζ itself is a root of $h(X^p)$. Hence $f(X)$ divides $h(X^p)$, and we can write

$$h(X^p) = f(X)g(X).$$

Since f has integral coefficients and leading coefficient 1, we see that g has integral coefficients. Since $a^p \equiv a \pmod{p}$ for any integer a, we conclude that

$$h(X^p) \equiv h(X)^p \pmod{p},$$

and hence

$$h(X)^p \equiv f(X)g(X) \pmod{p}.$$

In particular, if we denote by $\bar{f}$ and $\bar{h}$ the polynomials in $\mathbf{Z}/p\mathbf{Z}$ obtained by reducing f and h respectively mod p, we see that $\bar{f}$ and $\bar{h}$ are not relatively prime, i.e. have a factor in common. But $X^n - \bar{1} = \bar{f}(X)\bar{h}(X)$, and hence $X^n - \bar{1}$ has multiple roots. This is impossible, as one sees by taking the derivative, and our theorem is proved.

Corollary 3.2. *If n, m are relative prime integers $\geqq 1$, then*

$$\mathbf{Q}(\zeta_n) \cap \mathbf{Q}(\zeta_m) = \mathbf{Q}.$$

Proof. We note that ζ_n and ζ_m are both contained in $\mathbf{Q}(\zeta_{mn})$ since ζ_{mn}^n is a primitive m-th root of unity. Furthermore, $\zeta_m\zeta_n$ is a primitive mn-th root of unity. Hence

$$\mathbf{Q}(\zeta_n)\mathbf{Q}(\zeta_m) = \mathbf{Q}(\zeta_{mn}).$$

Our assertion follows from the multiplicativity $\varphi(mn) = \varphi(m)\varphi(n)$.

Suppose that n is a prime number p (having nothing to do with the characteristic). Then

$$X^p - 1 = (X - 1)(X^{p-1} + \cdots + 1).$$

Any primitive p-th root of unity is a root of the second factor on the right of this equation. Since there are exactly $p - 1$ primitive p-th roots of unity, we conclude that these roots are precisely the roots of

$$X^{p-1} + \cdots + 1.$$

We saw in Chapter IV, §3 that this polynomial could be transformed into an Eisenstein polynomial over the rationals. This gives another proof that $[\mathbf{Q}(\zeta_p)\colon \mathbf{Q}] = p - 1$.

We investigate more closely the factorization of $X^n - 1$, and suppose that we are in characteristic 0 for simplicity.

We have

$$X^n - 1 = \prod_{\zeta} (X - \zeta),$$

where the product is taken over all n-th roots of unity. Collect together all terms belonging to roots of unity having the same period. Let

$$\Phi_d(X) = \prod_{\text{period } \zeta = d} (X - \zeta)$$

Then

$$X^n - 1 = \prod_{d|n} \Phi_d(X).$$

We see that $\Phi_1(X) = X - 1$, and that

$$\Phi_n(X) = \frac{X^n - 1}{\prod\limits_{\substack{d|n \\ d<n}} \Phi_d(X)}.$$

From this we can compute $\Phi(X)$ recursively, and we see that $\Phi_n(X)$ is a polynomial in $\mathbf{Q}[X]$ because we divide recursively by polynomials having coefficients in $\mathbf{Q}$. All our polynomials have leading coefficient 1, so that in fact $\Phi_n(X)$ has *integer coefficients* by Theorem 1.1 of Chapter IV. Thus our construction is essentially universal and would hold over any field (whose characteristic does not divide n).

We call $\Phi_n(X)$ the n-th **cyclotomic polynomial**.

The roots of Φ_n are precisely the primitive n-th roots of unity, and hence

$$\deg \Phi_n = \varphi(n).$$

From Theorem 3.1 we conclude that Φ_n is irreducible over $\mathbf{Q}$, and hence

$$\Phi_n(X) = \operatorname{Irr}(\zeta_n, \mathbf{Q}, X).$$

We leave the proofs of the following recursion formulas as exercises:

1. If p is a prime number, then

$$\Phi_p(X) = X^{p-1} + X^{p-2} + \cdots + 1,$$

and for an integer $r \geqq 1$,

$$\Phi_{p^r}(X) = \Phi_p(X^{p^{r-1}}).$$

2. Let $n = p_1^{r_1} \cdots p_s^{r_s}$ be a positive integer with its prime factorization. Then

$$\Phi_n(X) = \Phi_{p_1 \cdots p_s}(X^{p_1^{r_1-1} \cdots p_s^{r_s-1}}).$$

3. If n is odd > 1, then $\Phi_{2n}(X) = \Phi_n(-X)$.

4. If p is a prime number, not dividing n, then

$$\Phi_{pn}(X) = \frac{\Phi_n(X^p)}{\Phi_n(X)}.$$

On the other hand, if $p \mid n$, then $\Phi_{pn}(X) = \Phi_n(X^p)$.

5. We have

$$\Phi_n(X) = \prod_{d \mid n} (X^{n/d} - 1)^{\mu(d)}.$$

As usual, μ is the Möbius function:

$$\mu(n) = \begin{cases} 0 & \text{if } n \text{ is divisible by } p^2 \text{ for some prime } p, \\ (-1)^r & \text{if } n = p_1 \cdots p_r \text{ is a product of distinct primes,} \\ 1 & \text{if } n = 1. \end{cases}$$

As an exercise, show that

$$\sum_{d \mid n} \mu(d) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if } n > 1. \end{cases}$$

Example. In light of Exercise 21 of Chapter V, we note that the association $n \mapsto \Phi_n(X)$ can be viewed as a function from the positive integers into the multiplicative group of non-zero rational functions. The multiplication formula $X^n - 1 = \prod \Phi_d(X)$ can therefore be inverted by the general formalism of convolutions. Computations of a number of cyclotomic polynomials show that for low values of n, they have coefficients equal to 0 or ± 1. However, I am indebted to Keith Conrad for bringing to my attention an extensive literature on the subject, starting with Bang in 1895. I include only the first and last items:

A. S. BANG, Om Ligningen $\Phi_m(X) = 0$, *Nyt Tidsskrift for Matematik* **(B) 6** (1895), pp. 6–12

H. L. MONTGOMERY and R. C. VAUGHN, The order of magnitude of the m-th coefficients of cyclotomic polynomials, *Glasgow Math. J.* **27** (1985), pp. 143–159

In particular, if $\Phi_n(X) = \Sigma a_{nj}X^j$, define $L(j) = \log \max_n |a_{nj}|$. Then Montgomery and Vaughn prove that

$$\frac{j^{1/2}}{(\log j)^{1/4}} \ll L(j) \ll \frac{j^{1/2}}{(\log j)^{1/4}}$$

where the sign $\ll$ means that the left-hand side is at most a positive constant times the right-hand side for $j \to \infty$. Bang also points out that $\Phi_{105}(X)$ is a cyclotomic polynomial of smallest degree having coefficients $\neq 0$ or ± 1: the coefficient of X^7 and X^{41} is -2 (all others are 0 or ± 1).

If ζ is an n-th root of unity and $\zeta \neq 1$, then

$$\frac{1 - \zeta^n}{1 - \zeta} = 1 + \zeta + \cdots + \zeta^{n-1} = 0.$$

This is trivial, but useful.

Let $\mathbf{F}_q$ be the finite field with q elements, q equal to a power of the odd prime number p. Then $\mathbf{F}_q^*$ has $q - 1$ elements and is a cyclic group. Hence we have the index

$$(\mathbf{F}_q^* : \mathbf{F}_q^{*2}) = 2.$$

If v is a non-zero integer not divisible by p, let

$$\left(\frac{v}{p}\right) = \begin{cases} 1 & \text{if } v \equiv x^2 \pmod{p} \text{ for some } x, \\ -1 & \text{if } v \not\equiv x^2 \pmod{p} \text{ for all } x. \end{cases}$$

This is known as the **quadratic symbol**, and depends only on the residue class of v mod p.

From our preceding remark, we see that there are as many quadratic residues as there are non-residues mod p.

Theorem 3.3. *Let ζ be a primitive p-th root of unity, and let*

$$S = \sum_v \left(\frac{v}{p}\right)\zeta^v,$$

the sum being taken over non-zero residue classes mod p. Then

$$S^2 = \left(\frac{-1}{p}\right)p.$$

Every quadratic extension of $\mathbf{Q}$ is contained in a cyclotomic extension.

Proof. The last statement follows at once from the explicit expression of $\pm p$ as a square in $\mathbf{Q}(\zeta)$, because the square root of an integer is contained in the

field obtained by adjoining the square root of the prime factors in its factorization, and also $\sqrt{-1}$. Furthermore, for the prime 2, we have $(1+i)^2 = 2i$. We now prove our assertion concerning S^2. We have

$$S^2 = \sum_{v,\mu} \left(\frac{v}{p}\right)\left(\frac{\mu}{p}\right)\zeta^{v+\mu} = \sum_{v,\mu} \left(\frac{v\mu}{p}\right)\zeta^{v+\mu}.$$

As v ranges over non-zero residue classes, so does $v\mu$ for any fixed μ, and hence replacing v by $v\mu$ yields

$$\begin{aligned} S^2 &= \sum_{v,\mu} \left(\frac{v\mu^2}{p}\right)\zeta^{\mu(v+1)} = \sum_{v,\mu} \left(\frac{v}{p}\right)\zeta^{\mu(v+1)} \\ &= \sum_{\mu} \left(\frac{-1}{p}\right)\zeta^0 + \sum_{v \neq -1} \left(\frac{v}{p}\right) \sum_{\mu} \zeta^{\mu(v+1)}. \end{aligned}$$

But $1 + \zeta + \cdots + \zeta^{p-1} = 0$, and the sum on the right over μ consequently yields -1. Hence

$$\begin{aligned} S^2 &= \left(\frac{-1}{p}\right)(p-1) + (-1)\sum_{v \neq -1} \left(\frac{v}{p}\right) \\ &= p\left(\frac{-1}{p}\right) - \sum_{v} \left(\frac{v}{p}\right) \\ &= p\left(\frac{-1}{p}\right), \end{aligned}$$

as desired.

We see that $\mathbf{Q}(\sqrt{p})$ is contained in $\mathbf{Q}(\zeta, \sqrt{-1})$ or $\mathbf{Q}(\zeta)$, depending on the sign of the quadratic symbol with -1. An extension of a field is said to be **cyclotomic** if it is contained in a field obtained by adjoining roots of unity. We have shown above that quadratic extensions of $\mathbf{Q}$ are cyclotomic. A theorem of Kronecker asserts that every abelian extension of $\mathbf{Q}$ is cyclotomic, but the proof needs techniques which cannot be covered in this book.

§4. LINEAR INDEPENDENCE OF CHARACTERS

Let G be a monoid and K a field. By a **character** of G in K (in this chapter), we shall mean a homomorphism

$$\chi: G \to K^*$$

of G into the multiplicative group of K. The **trivial character** is the homo-

morphism taking the constant value 1. Functions $f_i: G \to K$ are called **linearly independent** over K if whenever we have a relation

$$a_1 f_1 + \cdots + a_n f_n = 0$$

with $a_i \in K$, then all $a_i = 0$.

Examples. Characters will occur in various contexts in this book. First, the various conjugate embeddings of an extension field in an algebraic closure can be viewed as characters. These are the characters which most concern us in this chapter. Second, we shall meet characters in Chapter XVIII, when we shall extend the next theorem to a more general kind of character in connection with group representations.

Next, one meets characters in analysis. For instance, given an integer m, the function $f: \mathbf{R}/\mathbf{Z} \to \mathbf{C}^*$ such that $f(x) = e^{2\pi i m x}$ is a character on $\mathbf{R}/\mathbf{Z}$. It can be shown that all continuous homomorphisms of $\mathbf{R}/\mathbf{Z}$ into $\mathbf{C}^*$ are of this type. Similarly, given a real number y, the function $x \mapsto e^{2\pi i x y}$ is a continuous character on $\mathbf{R}$, and it is shown in Fourier analysis that all continuous characters of absolute value 1 on $\mathbf{R}$ are of this type.

Further, let X be a compact space and let R be the ring of continuous complex-valued functions on X. Let R^* be the group of units of R. Then given $x \in X$ the evaluation map $f \mapsto f(x)$ is a character of R^* into $\mathbf{C}^*$. (Actually, this evaluation map is a ring homomorphism of R onto $\mathbf{C}$.)

Artin found a neat way of expressing a linear independence property which covers all these cases, as well as others, in the following theorem [Ar 44].

Theorem 4.1. (Artin). *Let G be a monoid and K a field. Let $\chi_1, \ldots, \chi_n$ be distinct characters of G in K. Then they are linearly independent over K.*

Proof. One character is obviously linearly independent. Suppose that we have a relation

$$a_1 \chi_1 + \cdots + a_n \chi_n = 0$$

with $a_i \in K$, not all 0. Take such a relation with n as small as possible. Then $n \geqq 2$, and no a_i is equal to 0. Since χ_1, χ_2 are distinct, there exists $z \in G$ such that $\chi_1(z) \neq \chi_2(z)$. For all $x \in G$ we have

$$a_1 \chi_1(xz) + \cdots + a_n \chi_n(xz) = 0,$$

and since χ_i is a character,

$$a_1 \chi_1(z)\chi_1 + \cdots + a_n \chi_n(z)\chi_n = 0.$$

Divide by $\chi_1(z)$ and subtract from our first relation. The term $a_1\chi_1$ cancels, and we get a relation

$$\left(a_2 \frac{\chi_2(z)}{\chi_1(z)} - a_2\right)\chi_2 + \cdots = 0.$$

The first coefficient is not 0, and this is a relation of smaller length than our first relation, contradiction.

As an application of Artin's theorem, one can consider the case when K is a finite normal extension of a field k, and when the characters are distinct automorphisms $\sigma_1, \ldots, \sigma_n$ of K over k, viewed as homomorphisms of K^* into K^*. This special case had already been considered by Dedekind, who, however, expressed the theorem in a somewhat different way, considering the determinant constructed from $\sigma_i \omega_j$ where ω_j is a suitable set of elements of K, and proving in a more complicated way the fact that this determinant is not 0. The formulation given above and its particularly elegant proof are due to Artin.

As another application, we have:

Corollary 4.2. *Let $\alpha_1, \ldots, \alpha_n$ be distinct non-zero elements of a field K. If $a_1, \ldots, a_n$ are elements of K such that for all integers $v \geqq 0$ we have*

$$a_1 \alpha_1^v + \cdots + a_n \alpha_n^v = 0$$

then $a_i = 0$ for all i.

Proof. We apply the theorem to the distinct homomorphisms

$$v \mapsto \alpha_i^v$$

of $\mathbf{Z}_{\geqq 0}$ into K^*.

Another interesting application will be given as an exercise (relative invariants).

§5. THE NORM AND TRACE

Let E be a finite extension of k. Let $[E:k]_s = r$, and let

$$p^\mu = [E:k]_i$$

if the characteristic is $p > 0$, and 1 otherwise. Let $\sigma_1, \ldots, \sigma_r$ be the distinct embeddings of E in an algebraic closure k^a of k. If α is an element of E, we define its **norm** from E to k to be

$$N_{E/k}(\alpha) = N_k^E(\alpha) = \prod_{v=1}^{r} \sigma_v \alpha^{p^\mu} = \left(\prod_{v=1}^{r} \sigma_v \alpha \right)^{[E:k]_i}.$$

Similarly, we define the **trace**

$$\mathrm{Tr}_{E/k}(\alpha) = \mathrm{Tr}_k^E(\alpha) = [E:k]_i \sum_{v=1}^{r} \sigma_v \alpha.$$

The trace is equal to 0 if $[E:k]_i > 1$, in other words, if E/k is not separable.

Thus if E is separable over k, we have

$$N_k^E(\alpha) = \prod_\sigma \sigma\alpha$$

where the product is taken over the distinct embeddings of E in k^a over k.

Similarly, if E/k is separable, then

$$\mathrm{Tr}_k^E(\alpha) = \sum_\sigma \sigma\alpha.$$

Theorem 5.1. *Let E/k be a finite extension. Then the norm N_k^E is a multiplicative homomorphism of E^* into k^* and the trace is an additive homomorphism of E into k. If $E \supset F \supset k$ is a tower of fields, then the two maps are transitive, in other words,*

$$N_k^E = N_k^F \circ N_F^E \quad \text{and} \quad \mathrm{Tr}_k^E = \mathrm{Tr}_k^F \circ \mathrm{Tr}_F^E.$$

If $E = k(\alpha)$, and $f(X) = \mathrm{Irr}(\alpha, k, X) = X^n + a_{n-1}X^{n-1} + \cdots + a_0$, then

$$N_k^{k(\alpha)}(\alpha) = (-1)^n a_0 \quad \text{and} \quad \mathrm{Tr}_k^{k(\alpha)}(\alpha) = -a_{n-1}.$$

Proof. For the first assertion, we note that α^{p^μ} is separable over k if $p^\mu = [E:k]_i$. On the other hand, the product

$$\prod_{v=1}^{r} \sigma_v \alpha^{p^\mu}$$

is left fixed under any isomorphism into k^a because applying such an isomorphism simply permutes the factors. Hence this product must lie in k since α^{p^μ} is separable over k. A similar reasoning applies to the trace.

For the second assertion, let $\{\tau_j\}$ be the family of distinct embeddings of F into k^a over k. Extend each τ_j to an automorphism of k^a, and denote this extension by τ_j also. Let $\{\sigma_i\}$ be the family of embeddings of E in k^a over F. (Without loss of generality, we may assume that $E \subset k^a$.) If σ is an embedding of E over k in k^a, then for some j, $\tau_j^{-1}\sigma$ leaves F fixed, and hence $\tau_j^{-1}\sigma = \sigma_i$ for some i. Hence $\sigma = \tau_j\sigma_i$ and consequently the family $\{\tau_j\sigma_i\}$ gives all distinct embeddings of E into k^a over k. Since the inseparability degree is multiplicative in towers, our assertion concerning the transitivity of the norm and trace is obvious, because we have already shown that N_F^E maps E into F, and similarly for the trace.

Suppose now that $E = k(\alpha)$. We have

$$f(X) = ((X - \alpha_1) \cdots (X - \alpha_r))^{[E:k]_i}$$

if $\alpha_1, \ldots, \alpha_r$ are the distinct roots of f. Looking at the constant term of f gives us the expression for the norm, and looking at the next to highest term gives us the expression for the trace.

We observe that the trace is a k-linear map of E into k, namely

$$\mathrm{Tr}_k^E(c\alpha) = c\,\mathrm{Tr}_k^E(\alpha)$$

for all $\alpha \in E$ and $c \in k$. This is clear since c is fixed under every embedding of E over k. Thus the trace is a k-linear functional of E into k. For simplicity, we write $\mathrm{Tr} = \mathrm{Tr}_k^E$.

Theorem 5.2. *Let E be a finite separable extension of k. Then $\mathrm{Tr}: E \to k$ is a non-zero functional. The map*

$$(x, y) \mapsto \mathrm{Tr}(xy)$$

of $E \times E \to k$ is bilinear, and identifies E with its dual space.

Proof. That Tr is non-zero follows from the theorem on linear independence of characters. For each $x \in E$, the map

$$\mathrm{Tr}_x : E \to k$$

such that $\mathrm{Tr}_x(y) = \mathrm{Tr}(xy)$ is obviously a k-linear map, and the map

$$x \mapsto \mathrm{Tr}_x$$

is a k-homomorphism of E into its dual space $E^\vee$. (We don't write E^* for the dual space because we use the star to denote the multiplicative group of E.) If Tr_x is the zero map, then $\mathrm{Tr}(xE) = 0$. If $x \neq 0$ then $xE = E$. Hence the kernel of $x \mapsto \mathrm{Tr}_x$ is 0. Hence we get an injective homomorphism of E into the dual space $E^\vee$. Since these spaces have the same finite dimension, it follows that we get an isomorphism. This proves our theorem.

Corollary 5.3. *Let $\omega_1, \ldots, \omega_n$ be a basis of E over k. Then there exists a basis $\omega_1', \ldots, \omega_n'$ of E over k such that $\mathrm{Tr}(\omega_i \omega_j') = \delta_{ij}$.*

Proof. The basis $\omega_1', \ldots, \omega_n'$ is none other than the dual basis which we defined when we considered the dual space of an arbitrary vector space.

Corollary 5.4. *Let E be a finite separable extension of k, and let $\sigma_1, \ldots, \sigma_n$ be the distinct embeddings of E into k^a over k. Let $w_1, \ldots, w_n$ be elements of E. Then the vectors*

$$\xi_1 = (\sigma_1 w_1, \ldots, \sigma_1 w_n),$$
$$\cdots$$
$$\xi_n = (\sigma_n w_1, \ldots, \sigma_n w_n)$$

are linearly independent over E if $w_1, \ldots, w_n$ form a basis of E over k.

Proof. Assume that $w_1, \ldots, w_n$ form a basis of E/k. Let $\alpha_1, \ldots, \alpha_n$ be elements of E such that

$$\alpha_1 \xi_1 + \cdots + \alpha_n \xi_n = 0.$$

Then we see that

$$\alpha_1 \sigma_1 + \cdots + \alpha_n \sigma_n$$