Math 250A: Groups, rings, and fields.
H. W. Lenstra jr.

1. Prerequisites

This section consists of an enumeration of terms from elementary set theory and algebra. You are supposed to be familiar with their definitions and basic properties.

Set theory. Sets, subsets, the empty set $\emptyset$, operations on sets (union, intersection, product), maps, composition of maps, injective maps, surjective maps, bijective maps, the identity map 1_X of a set X, inverses of maps. Relations, equivalence relations, equivalence classes, partial and total orderings, the cardinality $\#X$ of a set X. The principle of mathematical induction. Zorn's lemma will be assumed in a number of exercises. Later in the course the terminology and a few basic results from *point set topology* may come in useful.

Group theory. Groups, multiplicative and additive notation, the unit element 1 (or the zero element 0), abelian groups, cyclic groups, the order of a group or of an element, Fermat's little theorem, products of groups, subgroups, generators for subgroups, left cosets aH, right cosets, the coset spaces G/H and $H\backslash G$, the index $(G : H)$, the theorem of Lagrange, group homomorphisms, isomorphisms, automorphisms, normal subgroups, the factor group G/N and the canonical map $G \to G/N$, homomorphism theorems, the Jordan-Hölder theorem (see Exercise 1.4), the commutator subgroup $[G, G]$, the center $Z(G)$ (see Exercise 1.12), the group $\operatorname{Aut} G$ of automorphisms of G, inner automorphisms.

Examples of groups: the group $\operatorname{Sym} X$ of permutations of a set X, the symmetric group $S_n = \operatorname{Sym}\{1, 2, \ldots, n\}$, cycles of permutations, even and odd permutations, the alternating group A_n, the dihedral group $D_n = \langle (1\ 2\ \ldots\ n), (1\ n{-}1)(2\ n{-}2)\ldots \rangle$, the Klein four group V_4, the quaternion group $Q_8 = \{\pm 1, \pm i, \pm j, \pm ij\}$ (with $ii = jj = -1$, $ji = -ij$) of order 8, additive groups of rings, the group $\operatorname{Gl}(n, R)$ of invertible $n \times n$-matrices over a ring R.

Occasionally the structure theorem of *finite abelian groups* and *finitely generated abelian groups* will be assumed known.

Ring theory. Rings, subrings, homomorphisms. We adopt the convention that the existence of a unit element 1 is part of the definition of a ring. Likewise, subrings of a ring R are required to contain the unit element of R, and ring homomorphisms are required to map 1 to 1. Products of rings, zero-divisors, units, the group R^* of units of a ring R. Ideals, operations on ideals (sum, intersection, product), generators for ideals, principal ideals, the factor ring $R/\mathfrak{a}$, congruence modulo an ideal. Domains, prime ideals, maximal ideals. Principal ideal rings, unique factorization domains.

Examples: the ring $\mathbf{Z}$ of integers, the ring $\mathbf{Z}/n\mathbf{Z}$ of integers modulo n, the zero ring, polynomial rings $R[X_1, \ldots, X_n]$, the endomorphism ring $\operatorname{End} A$ of an abelian group A (see Exercise 1.23), the ring $M(n, R)$ of $n \times n$-matrices over a ring R, the division ring (or skew field) $\mathbf{H}$ of quaternions, the group ring $R[G]$ of a group G over a ring R (see Exercise 1.25), the ring $R[[X]]$ of formal power series over a ring R. (see Exercise 1.26).

Field theory. Fields, the field $Q(R)$ of fractions of a domain R, the prime field, the characteristic $\operatorname{char} k$ of a field k, algebraic and transcendental elements over a field, algebraic and finite extensions, the degree $[l:k]$, the irreducible (or minimum) polynomial f_k^α of an element α algebraic over k, the isomorphism $k(\alpha) \cong k[X]/f_k^\alpha k[X]$, splitting fields.

Examples: the fields $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{C}$ of rational, real, and complex numbers, respectively, the field $\mathbf{F}_p = \mathbf{Z}/p\mathbf{Z}$ (for p prime), finite fields $\mathbf{F}_q$ (for q a prime power), the field $k(X_1, \ldots, X_n)$ of rational functions in n indeterminates over a field k.

1. Exercises

1.1 (Left axioms). Let G be a set, $1 \in G$ an element, $i\colon G \to G$ a map, and $m\colon G \times G \to G$ another map; instead of $m(x, y)$ we write xy, for $x, y \in G$.

(a) Suppose that for all $x, y, z \in G$ one has

$$1x = x, \quad i(x)x = 1, \quad (xy)z = x(yz).$$

Prove that G is a group with multiplication m.

(b) Suppose that for all $x, y, z \in G$ one has

$$x1 = x, \quad i(x)x = 1, \quad (xy)z = x(yz).$$

Does it follow that G is a group with multiplication m? Give a proof or a counterexample.

1.2. Let u, U, v, V be subgroups of a group G. Assume that u is a normal subgroup of U and that v is a normal subgroup of V. Prove that $u(U \cap v)$ is a normal subgroup of $u(U \cap V)$, that $(u \cap V)(U \cap v)$ is a normal subgroup of $U \cap V$, that $(u \cap V)v$ is a normal subgroup of $(U \cap V)v$, and that there are group isomorphisms

$$u(U \cap V)/\big(u(U \cap v)\big) \cong (U \cap V)/\big((u \cap V)(U \cap v)\big) \cong (U \cap V)v/\big((u \cap V)v\big).$$

1.3. A *normal tower* of a group G is a sequence $(u_i)_{i=0}^r$ of subgroups u_i of G for which r is a non-negative integer, each u_{i-1} is a normal subgroup of u_i (for $0 < i \le r$), and $u_0 = \{1\}$, $u_r = G$. A *refinement* of a normal tower $(u_i)_{i=0}^r$ is a normal tower $(w_j)_{j=0}^t$ with the property that there is an increasing map $f\colon \{0, 1, 2, \ldots, r\} \to \{0, 1, 2, \ldots, t\}$ such that for all i one has $u_i = w_{f(i)}$. Two normal towers $(u_i)_{i=0}^r$ and $(v_j)_{j=0}^s$ are called *equivalent* if there is a bijection $g\colon \{1, 2, \ldots, r\} \to \{1, 2, \ldots, s\}$ such that for all i one has $u_i/u_{i-1} \cong v_{g(i)}/v_{g(i)-1}$.

Prove the **Schreier refinement theorem**: any two normal towers of a group have equivalent refinements.

1.4. A group is called *simple* if it has exactly two normal subgroups (namely $\{1\}$ and the group itself). A *composition series* of a group is a normal tower $(u_i)_{i=0}^r$ for which every group u_i/u_{i-1} is simple.

(a) Prove the **Jordan-Hölder theorem**: any two composition series of a group are equivalent.

(b) Exhibit a group that does not have a composition series.

(c) Classify all simple abelian groups.

1.5. (a) Prove that, up to isomorphism, there are precisely 9 groups of order smaller than 8. Which are they? Formulate the theorems that you use.

(b) How many pairwise non-isomorphic groups of order 8 are there? Prove the correctness of your answer.

1.6. Let H, N be groups, and let $\psi\colon H \to \operatorname{Aut} N$ be a group homomorphism. Instead of $\psi(h)(x)$ we shall write ${}^h x$, for $h \in H$, $x \in N$.

(a) Prove that the set $N \times H$ with the operation

$$(x, h) \cdot (y, h') = (x \cdot {}^h y, hh')$$

is a group. This group is called the *semidirect product* of H and N (with respect to ψ), notation: $N \rtimes H$.

(b) Prove that N may be viewed as a normal subgroup of $N \rtimes H$, and that $(N \rtimes H)/N$ is isomorphic to H.

(c) Is the direct product a special case of the semidirect product?

1.7. (a) Let G be a group, let H be a subgroup of G, and let N be a normal subgroup of G. Suppose that the composition of the inclusion $H \subset G$ with the natural map $G \to G/N$ is an isomorphism $H \xrightarrow{\sim} G/N$. Prove that there is a group homomorphism $\psi\colon H \to \operatorname{Aut} N$ such that the map $N \rtimes H \to G$ sending (x, h) to xh is a group isomorphism.

(b) Let N_1 and N_2 be two normal subgroups of a group G with $N_1 \cap N_2 = \{1\}$. Prove that for all $x \in N_1$, $y \in N_2$ one has $xy = yx$.

1.8. Prove that S_4 is isomorphic to $V_4 \rtimes S_3$ with respect to some *isomorphism* $\psi\colon S_3 \to \operatorname{Aut} V_4$.

1.9. A subgroup H of a group G is called *characteristic* (in G) if for all automorphisms φ of G one has $\varphi H = H$.

(a) Prove that any characteristic subgroup of a group is normal.

(b) Give an example of a group and a normal subgroup that is not characteristic.

1.10. (a) Let H be a subgroup of a group G, and let J be a characteristic subgroup of H. Prove: if H is normal in G then J is normal in G, and if H is characteristic in G then J is characteristic in G.

(b) If H is a normal subgroup of a group G, and J is a normal subgroup of H, does it follow that J is normal in G? Give a proof or a counterexample.

1.11. (a) Prove that there is a non-abelian group G of order 8 with the property that every subgroup of G is normal.

(b) Does there exist a group that has *exactly one* subgroup that is not normal? Give an example, or prove that no such group exists.

1.12. Let G be a group. By $Z(G)$ we denote the *center* of G, i. e.:

$$Z(G) = \{x \in G : \text{for all } y \in G \text{ one has } xy = yx\}.$$

(a) Prove that the commutator subgroup $[G, G]$ and the center $Z(G)$ of G are characteristic subgroups of G.

(b) Prove that every subgroup of G that contains $[G, G]$ or is contained in $Z(G)$ is normal in G.

(c) Prove: the group $G/Z(G)$ is cyclic if and only if G is abelian.

1.13. Let G be a group.

(a) Prove that the group of inner automorphisms of G is a normal subgroup of $\operatorname{Aut} G$ that is isomorphic to $G/Z(G)$.

(b) Suppose that $Z(G) = \{1\}$. Prove that $Z(\operatorname{Aut} G) = \{1\}$.

1.14. Let G be a group. An *anti-automorphism* of G is a bijective map $\varphi\colon G \to G$ such that for all x, $y \in G$ one has $\varphi(xy) = \varphi(y)\varphi(x)$. Write $\operatorname{Ant} G$ for the set of anti-automorphisms of G.

(a) Prove that $\operatorname{Aut} G \cup \operatorname{Ant} G$ is a group, the group operation being composition of maps.

(b) Suppose that G is non-abelian. Prove that $\operatorname{Aut} G \cup \operatorname{Ant} G$ is isomorphic to the product of $\operatorname{Aut} G$ and a group of order 2.

1.15. Let G be a group with the property that $x = x^{-1}$ for all $x \in G$.

(a) Prove that G is abelian.

(b) Call a subset $S \subset G$ *independent* if for every finite non-empty subset T of S one has $\prod_{t \in T} t \neq 1$. Use Zorn's lemma to prove that G contains an independent subset S that is *maximal* in the sense that the only independent subset U of G with $S \subset U \subset G$ is S itself.

(c) Suppose that S is a maximal independent subset of G. Prove that for every $x \in G$ there is a unique finite subset $S(x)$ of S with the property that $x = \prod_{s \in S(x)} s$. Prove also that for x, $y \in G$ one has $S(xy) = S(x) \bigtriangleup S(y)$; here the *symmetric difference* $A \bigtriangleup B$ of two subsets A and B of a set C is defined by $A \bigtriangleup B = \{x \in C : x \in A \text{ and } x \notin B\} \cup \{x \in C : x \notin A \text{ and } x \in B\}$.

1.16. Let G be a group. Prove: $\operatorname{Aut} G = \{1\}$ if and only if G has order 1 or 2.

1.17 (Goursat's lemma). Let G_1 and G_2 be groups, and let $H \subset G_1 \times G_2$ be a subgroup. Prove that there are subgroups H_1, H_2 of G_1, G_2 (respectively), normal subgroups N_1, N_2 of H_1, H_2 (respectively), and a group isomorphsim $\psi\colon H_1/N_1 \xrightarrow{\sim} H_2/N_2$, such that H is the 'graph' of ψ in the sense that

$$H = \{(x, y) : x \in H_1,\ y \in H_2,\ \psi(xN_1) = yN_2\}.$$

Are H_1, H_2, N_1, N_2, and ψ uniquely determined by H?

1.18. (a) Let G be a group, and let $x \in G$ have order 2. Prove: $\langle x \rangle$ is normal in G if and only if $x \in Z(G)$.

(b) Let G be a finite group of odd order, and let $x \in G$ have order 17. Prove: $\langle x \rangle$ is normal in G if and only if $x \in Z(G)$.

1.19. (a) Let G be a finite abelian group of squarefree order (i. e., the order is not divisible by the square of a prime number). Prove that G is cyclic.

(b) Let G be a cyclic group. Prove that $\operatorname{Aut} G$ is abelian.

1.20. In this problem we write $G' = [G, G]$ for a group G, and $G'' = (G')'$, $G''' = (G'')'$.

Let G be a finite group of squarefree order. Prove that $G'' = G'''$. (*Hint*: study a naturally defined map $G \to \operatorname{Aut}(G''/G''')$.) *Note*: one can actually show that for any finite group G of squarefree order one has $G'' = \{1\}$; see Exercise 3.27.

1.21. Let R be a commutative ring. Two ideals $\mathfrak{a}$, $\mathfrak{b}$ of R are said to be *coprime* if $\mathfrak{a}+\mathfrak{b} = R$.

(a) Prove: two ideals $\mathfrak{a}$, $\mathfrak{b}$ are coprime if and only if there exist $a \in \mathfrak{a}$ and $b \in \mathfrak{b}$ with $a + b = 1$, and if and only if the map $R \to (R/\mathfrak{a}) \times (R/\mathfrak{b})$ sending x to $(x + \mathfrak{a}, x + \mathfrak{b})$ is surjective.

(b) Let $\mathfrak{a}_1$, $\mathfrak{a}_2$, ..., $\mathfrak{a}_t$, $\mathfrak{b}_1$, $\mathfrak{b}_2$, ..., $\mathfrak{b}_u$ be ideals of R, and suppose that $\mathfrak{a}_i$ and $\mathfrak{b}_j$ are coprime for every i, j. Prove that the ideals $\mathfrak{a}_1\mathfrak{a}_2 \cdots \mathfrak{a}_t$ and $\mathfrak{b}_1\mathfrak{b}_2 \cdots \mathfrak{b}_u$ are coprime.

1.22. Let R be a commutative ring.

(a) Prove the **Chinese remainder theorem**: if $\mathfrak{a}$, $\mathfrak{b}$ are coprime ideals of R, then the map $R \to (R/\mathfrak{a}) \times (R/\mathfrak{b})$ from Exercise 1.21(a) induces a ring isomorphism $R/(\mathfrak{a}\mathfrak{b}) \to (R/\mathfrak{a}) \times (R/\mathfrak{b})$.

(b) Formulate and prove a version of the Chinese remainder theorem for pairwise coprime ideals $\mathfrak{a}_1$, $\mathfrak{a}_2$, ..., $\mathfrak{a}_t$ of R.

1.23. Let A be an additively written abelian group. An *endomorphism* of A is a group homomorphism $A \to A$. Denote by $\operatorname{End} A$ the set of all endomorphisms of A.

(a) Prove that $\operatorname{End} A$ is a ring, with sum and product defined as follows: $(f+g)(x) = f(x) + g(x)$, $(fg)(x) = f\big(g(x)\big)$, for f, $g \in \operatorname{End} A$, $x \in A$. What goes wrong if A is not abelian?

(b) Prove that $\operatorname{End} V_4$ is isomorphic to the ring of 2×2-matrices over the field $\mathbf{F}_2 = \mathbf{Z}/2\mathbf{Z}$.

1.24. (a) Show that each ring R is isomorphic to a subring of the ring $\operatorname{End} R^+$ of endomorphisms of the additive group R^+ of R.

(b) Let R be a ring of which the additive group is cyclic. Prove that R, as a ring, is isomorphic to $\mathbf{Z}/n\mathbf{Z}$ for some non-negative integer n.

1.25. Let R be a ring and G a group. The *group ring* $R[G]$ of G over R is the set of all formal expressions $\sum_{g\in G} r_g g$, where $r_g \in R$ for all $g \in G$, and $r_g = 0$ for all but finitely many $g \in G$. Sum and product are defined by

$$\Big(\sum_{g\in G} r_g g\Big)+\Big(\sum_{g\in G} s_g g\Big) = \sum_{g\in G}(r_g+s_g)g, \qquad \Big(\sum_{g\in G} r_g g\Big)\cdot\Big(\sum_{g\in G} s_g g\Big) = \sum_{g\in G}\Big(\sum_{a,b\in G, ab=g} r_a s_b\Big)g.$$

Prove that this is indeed a ring, and give necessary and sufficient conditions in terms of R and G for $R[G]$ to be commutative.

1.26. Let R be a ring and X a symbol. The *ring* $R[[X]]$ *of formal power series* in X over R is the set of all formal expressions $\sum_{i=0}^{\infty} r_i X^i$, where $r_i \in R$ for all $i \geq 0$. Sum and product are defined by

$$\Big(\sum_{i=0}^{\infty} r_i X^i\Big) + \Big(\sum_{i=0}^{\infty} s_i X^i\Big) = \sum_{i=0}^{\infty}(r_i + s_i)X^i,$$

$$\Big(\sum_{i=0}^{\infty} r_i X^i\Big) \cdot \Big(\sum_{i=0}^{\infty} s_i X^i\Big) = \sum_{i=0}^{\infty}\Big(\sum_{j=0}^{i} r_j s_{i-j}\Big)X^i.$$

Prove that this is indeed a ring, and that the polynomial ring $R[X]$ can be viewed as a subring of $R[[X]]$.

1.27. Let R be a commutative ring.

(a) Use Zorn's lemma to prove: R has a maximal ideal if and only if R is not the zero ring.

(b) Prove that the union of all maximal ideals of R is equal to the set $R - R^*$ of non-units of R.

1.28. Let a *rng* be an additively written abelian group R equipped with an associative bilinear map $R \times R \to R$ that is multiplicatively written (bilinearity means $a(b+c) = ab+ac$ and $(a + b)c = ac + bc$ for all a, b, $c \in R$; so a rng differs from a ring only in that no multiplicative identity is required to exist). Commutativity, ideals, and maximality of ideals are defined for rngs as they are for rings.

Show that there exists a non-zero commutative rng that has no maximal ideal.

1.29. Let R be a commutative ring. The *Jacobson radical* $J(R)$ of R is defined to be the intersection of all maximal ideals of R. Prove that $J(R)$ equals the set of all $a \in R$ with the property that for all $r \in R$ the element $1 + ra$ is a unit of R.

1.30. Let R be a commutative ring, and let $S \subset R$ be a *multiplicative set*, i. e.: $1 \in S$, and $st \in S$ whenever $s, t \in S$.

(a) Define the relation $\sim$ on $R \times S$ by $(x, s) \sim (y, t)$ if and only if there exists $u \in S$ with $utx = usy$. Prove that $\sim$ is an equivalence relation.

(b) Denote by $S^{-1}R$ the set of equivalence classes of $\sim$, and the class of (r, s) by r/s. Prove that there is a unique commutative ring structure on $S^{-1}R$ with the following two properties: (i) the map $R \to S^{-1}R$ sending r to $r/1$ is a ring homomorphism; (ii) the product of r/s and $s/1$ is $r/1$, for each $(r, s) \in R \times S$.

(c) Prove that $S^{-1}R$ is the zero ring if and only if $0 \in S$.

1.31. Let R be a commutative ring, and let $S \subset R$ be a multiplicative set with $0 \notin S$. Prove that R has a prime ideal $\mathfrak{p}$ with $\mathfrak{p} \cap S = \emptyset$.

1.32. An element x of a ring is called *nilpotent* if there exists a positive integer n with $x^n = 0$. The *nilradical* $\sqrt{0}_R$ of a commutative ring R is the set of nilpotent elements in R.

Let R be a commutative ring. Prove that $\sqrt{0}_R$ is an ideal of R, and that it is equal to the intersection of all prime ideals of R.

1.33. An element e of a ring is called *idempotent* if $e^2 = e$. Let R be a commutative ring, and write $B(R)$ for the set of its idempotents.

(a) We call R *connected* if $R \neq \{0\}$ and there do not exist non-zero rings R_1 and R_2 such that $R \cong R_1 \times R_2$ (as rings). Prove that R is connected if and only if the cardinality of $B(R)$ equals 2.

(b) Prove that $B(R)$ is a ring with addition & defined by $e_1 \& e_2 = e_1 + e_2 - 2e_1e_2$ and the multiplication taken from R.

1.34. A *Boolean ring* is a ring A with the property that $B(A) = A$.

(a) Let A be a Boolean ring. Prove that A is commutative, and that there is a ring homomorphism $\mathbf{F}_2 \to A$.

(b) Prove that for every commutative ring R, the ring $B(R)$ defined in Exercise 1.33 is Boolean.

1.35. Let R be a commutative ring, and let $B(R)$ be as in Exercise 1.33.

(a) Prove that $(e_1 - e_2)^n = e_1 - e_2$ for all $e_1, e_2 \in B(R)$ and all positive odd integers n.

(b) Prove that for each positive integer n there exists a polynomial $f \in X^n\mathbf{Z}[X]$ with $f - 1 \in (X - 1)^n\mathbf{Z}[X]$; that any such f satisfies $f \equiv X \bmod (X^2 - X)\mathbf{Z}[X]$; and that there is a unique such f of degree smaller than $2n$. Find that unique f for $n = 2$.

(c) Let $\sqrt{0}_R$ denote the nilradical of R. Show that the natural map $R \to R/\sqrt{0}_R$ induces a bijection $B(R) \to B(R/\sqrt{0}_R)$.

1.36. Let X be a set, and let $P(X)$ be the set of all subsets of X. Define an addition and a multiplication on $P(X)$ by $A + B = A \triangle B$ (see Exercise 1.15(c)) and $A \cdot B = A \cap B$, for A, $B \subset X$. Prove that these operations make $P(X)$ into a ring, and that this ring is Boolean.

1.37. Let R be a Boolean ring.

(a) Prove: if $\mathfrak{a} \subset R$ is an ideal, then $\mathfrak{a}$ is maximal if and only if $\mathfrak{a}$ is prime, and if and only if $R/\mathfrak{a} \cong \mathbf{F}_2$.

(b) Let X be the set of prime ideals of R. Prove that R is isomorphic to a subring of the ring $P(X)$ defined in Exercise 1.36.

1.38. Let X be a set, and let $P(X)$ be as in Exercise 1.36. A *filter* on X is a non-empty subset $\mathcal{F} \subset P(X)$ with the properties (i) if $A \in \mathcal{F}$, $B \subset X$ are such that $A \subset B$, then $B \in \mathcal{F}$; and (ii) if A, $B \in \mathcal{F}$, then $A \cap B \in \mathcal{F}$. An *ultrafilter* on X is a filter $\mathcal{U}$ on X with $\emptyset \notin \mathcal{U}$ with the property that the only filters $\mathcal{F}$ on X with $\mathcal{U} \subset \mathcal{F}$ are $\mathcal{F} = \mathcal{U}$ and $\mathcal{F} = P(X)$.

(a) Let $\mathcal{F} \subset X$. Prove: $\mathcal{F}$ is a filter if and only if $\{A \subset X : X - A \in \mathcal{F}\}$ is an ideal of the ring $P(X)$ (here $X - A$ denotes the complement of A in X).

(b) Prove: every filter $\mathcal{F}$ on X with $\emptyset \notin \mathcal{F}$ is contained in an ultrafilter on X.

(c) Suppose that $\mathcal{U}$ is an ultrafilter on X. Prove that for any $A \subset X$ one has either $A \in \mathcal{U}$ or $X - A \in \mathcal{U}$, but not both.

(d) Suppose X is infinite. Prove that there is an ultrafilter on X that does not contain any finite set.

1.39. Let R be a finite ring. Prove that there are integers n, m with $n > m > 0$ such that for all $x \in R$ one has $x^n = x^m$.

1.40. Let T be a subring of a finite ring R. Prove: $T^* = T \cap R^*$. Give a counterxample when the finiteness condition is omitted.

1.41. Let R be a finite commutative ring, and for a maximal ideal $\mathfrak{m} \subset R$ let the *norm* $\mathfrak{N}\mathfrak{m}$ be the cardinality of the residue class field $R/\mathfrak{m}$ of $\mathfrak{m}$. Prove that

$$\#R^* = \#R \cdot \prod_{\mathfrak{m}} \Bigl(1 - \frac{1}{\mathfrak{N}\mathfrak{m}}\Bigr),$$

with $\mathfrak{m}$ ranging over the set of maximal ideals of R.

1.42. Let R be a commutative ring, and let $T \subset R$ be a subring for which the index $(R : T)$ of additive groups is finite. Define $\mathfrak{a} = \{x \in R : xR \subset T\}$.

(a) Prove that $\mathfrak{a}$ is an R-ideal that is contained in T.
(b) Prove that $R/\mathfrak{a}$ is finite.
(c) Compute $\mathfrak{a}$ in the case $R = \mathbf{Z}[(1+\sqrt{-3})/2]$ and $T = \mathbf{Z}[\sqrt{-3}]$.

2. Actions of groups

Let G be a group and X a set. An *action* of G on X is a map $G \times X \to X$, $(\sigma, x) \mapsto \sigma x$, satisfying the two axioms

$$1x = x, \qquad \sigma(\tau x) = (\sigma\tau)x$$

for all $x \in X$, σ, $\tau \in G$. If an action of G on X is given, then G is said to *act* on X, and X is called a *G-set*.

An action is also called an *operation*, a *permutation action*, or a *left action*. A *right action* of G on X is a map $G \times X \to X$, $(\sigma, x) \mapsto x\sigma$, satisfying $x1 = x$ and $(x\sigma)\tau = x(\sigma\tau)$ for all $x \in X$, σ, $\tau \in G$. Since there is an easy dictionary between right and left actions (see Exercise 2.1(a)), only one of the two needs to be considered.

If X is a G-set, then one readily checks that the map $f_\sigma\colon X \to X$ sending x to σx is bijective, with two-sided inverse $f_{\sigma^{-1}}$; also, the map $G \to \operatorname{Sym} X$ sending σ to f_σ is a group homomorphism. Conversely, any group homomorphism $\psi\colon G \to \operatorname{Sym} X$ gives rise to an action of G on X, by $\sigma x = \big(\psi(\sigma)\big)(x)$. Hence giving an action of G on X is equivalent to giving a group homomorphism $G \to \operatorname{Sym} X$.

If X and Y are G-sets, then a *G-map* or *map of G-sets* $X \to Y$ is a map $f\colon X \to Y$ satisfying $f(\sigma x) = \sigma(fx)$ for all $\sigma \in G$, $x \in X$. The composition of two G-maps (if defined) is a G-map, and the identity map from any G-set to itself is a G-map. A *G-isomorphism* is a bijective G-map, and a *G-automorphism* of a G-set X is a G-isomorphism from X to itself. Two G-sets X and Y are called (G-)isomorphic, notation $X \cong_G Y$, if there exists a G-isomorphism $X \to Y$.

Examples. (i) For any set X, the group $\operatorname{Sym} X$ and any of its subgroups acts on X in an obvious manner. If R is a ring and n is a non-negative integer, then $\mathrm{Gl}(n, R)$ acts on R^n: namely, let Ax for $A \in \mathrm{Gl}(n, R)$ and $x \in R^n$ be the usual product of a matrix and a vector; likewise, if m is another non-negative integer, then $\mathrm{Gl}(n, R)$ acts on the set of $n \times m$-matrices over R by matrix multiplication. If R is any ring, then R^* acts on R by the multiplication in the ring; in this example, *right* multiplication by any element of R is an R^*-map from R to itself.

(ii) Let G be any group. There are at least three naturally occurring actions $G \times G \to G$ of G on G itself. The first is left multiplication: $(\sigma, \tau) \to \sigma\tau$ (the product in the group). The second is right multiplication $(\sigma, \tau) \to \tau\sigma^{-1}$; note that the inverse of σ is taken, since otherwise we would have a *right* action (cf. Exercise 2.1(a)). The third action is by *conjugation*: $(\sigma, \tau) \to \sigma\tau\sigma^{-1}$. We write ${}^\sigma\tau = \sigma\tau\sigma^{-1}$, the exponent being on the *left*, so that the second axiom takes the shape ${}^\sigma({}^\tau x) = {}^{\sigma\tau}x$. For conjugation, many group theorists prefer the corresponding *right* action $(\sigma, \tau) \mapsto \sigma^{-1}\tau\sigma = \tau^\sigma$; the second axiom of a right action then takes the familiar shape $(x^\sigma)^\tau = x^{\sigma\tau}$. Likewise, G acts by conjugation on the set of subgroups H of G, by ${}^\sigma H = \sigma H \sigma^{-1}$.

(iii) Let G be a group, and let H be a subgroup of G. Then H acts on G by left multiplication, and in fact H acts in this way on any union of right cosets $Ha \subset G$ of H in G; for example, any line in a vector space that passes through the origin acts by translation on any line that is parallel to it.

(iv) Let again G be a group, H a subgroup of G, and let $G/H = \{aH : a \in G\}$ be the set of left cosets of H in G. Then G acts on G/H by $a(bH) = (ab)H$. One needs to check that this is well-defined, and that it is an action, but both are easy. The natural map $G \to G/H$ sending a to aH is a G-map.

(v) Let $n \in \mathbf{Z}$, $n \geq 3$. Then the maps σ, $\tau\colon \mathbf{C} \to \mathbf{C}$ defined by $\sigma z = e^{2\pi i/n}z$ and $\tau z = \bar{z}$ are bijective. In $\operatorname{Sym}\mathbf{C}$, they satisfy $\sigma^n = \tau^2 = 1_{\mathbf{C}}$ and $\tau\sigma\tau^{-1} = \sigma^{-1}$. One deduces that they generate a subgroup of $\operatorname{Sym}\mathbf{C}$ that is isomorphic to the dihedral group D_n. Thus D_n acts on $\mathbf{C}$.

Actions of subgroups. Let a group G act on a set X. For any subgroup H of G, an action of H on X is induced: just restrict the map $G \times X \to X$ to $H \times X$. More generally, if $f\colon K \to G$ is a group homomorphism of some group K to G, then an action of K on X is induced "via f": one defines σx to be $f(\sigma)x$, for $\sigma \in K$, $x \in X$. For example, the three actions of G on G that we defined under (ii) all come from the single action of $G \times G$ on G that is defined by $(\sigma, \tau)\rho = \sigma\rho\tau^{-1}$, via three embeddings $G \to G \times G$: the two "coordinate" embeddings $\sigma \mapsto (\sigma, 1)$ and $\sigma \mapsto (1, \sigma)$, and the "diagonal" embedding $\sigma \mapsto (\sigma, \sigma)$.

Actions on subsets. A subset Y of a G-set is called *stable* under the action of G, or a *G-subset*, if for all $\sigma \in G$ and $y \in Y$ one has $\sigma y \in Y$.

Trivial, free, and faithful actions. Any group G can be made to act on any set X by putting $\sigma x = x$ for all $\sigma \in G$ and $x \in X$; this is called the *trivial* action, and the corresponding G-set X is called trivial. An action of G on X (or the G-set X itself) is called *free* if for all $\sigma \in G$, $\sigma \neq 1$, and all $x \in X$ one has $\sigma x \neq x$; and it is called *faithful* if for all $\sigma \in G$, $\sigma \neq 1$, there exists $x \in X$ with $\sigma x \neq x$. For some obvious properties of these notions, see Exercise 2.2.

Fixed points. Let G be a group and X a G-set. A *fixed point* (or *fixpoint*) of X is an element $x \in X$ with the property that for all $\sigma \in G$ one has $\sigma x = x$. The set of all fixed points is denoted by X^G or, if there is danger of confusion with some other use of the exponential notation (see Exercise 2.4), by $\operatorname{Fix}_G X$. It is an example of a G-subset of X.

Orbits. Let G be a group and X a G-set. Two elements x, $y \in X$ are said to be *equivalent under* G if there exists $\sigma \in G$ with $\sigma x = y$; one readily checks that this is indeed an equivalence relation. The equivalence classes of this equivalence relation are called the *orbits* of X under G, or of the action of G on X. The orbit containing a given element

$x \in X$ is equal to $Gx = \{\sigma x : \sigma \in G\}$. Thus, X can be written as a disjoint union $\bigcup_x Gx$, where x ranges over a set of representatives for the orbits. Each of these orbits is a G-subset of X. For example, if a subgroup H of G acts on G by left multiplication, then the orbits are the right cosets Ha, $a \in G$, and $H\backslash G$ is the set of orbits of G under H. In general, the set of orbits of X under G should be denoted by $G\backslash X$, so as to be consistent with the notation $H\backslash G$; but one often sees the notation X/G, which should really be reserved for right actions. If one provides $G\backslash X$ with the trivial G-action, then the natural map $X \to G\backslash X$ that sends x to the orbit Gx is a G-map.

Stabilizers. Let G be a group, X a G-set, and $x \in X$. The *stabilizer* G_x of x in G is defined by

$$G_x = \{\sigma \in G : \sigma x = x\}.$$

(This is also called the *isotropy group* or *decomposition group* of x.) One readily checks that G_x is a subgroup of G. For example, if $X = G/H$ for some subgroup $H \subset G$, and $x = H \in X$, then we have $G_x = H$. For $\tau \in G$ one has

$$G_{\tau x} = \tau G_x \tau^{-1}.$$

To prove this, let $\sigma \in G$. Then: $\sigma \in G_{\tau x} \Leftrightarrow \sigma\tau x = \tau x \Leftrightarrow \tau^{-1}\sigma\tau x = x \Leftrightarrow \tau^{-1}\sigma\tau \in G_x \Leftrightarrow \sigma \in \tau G_x \tau^{-1}$.

The structure of orbits. Let G be a group acting on itself by left multiplication, let X be a G-set, and let $x \in X$. Then the map $G \to X$ sending σ to σx is a G-map. The image of this map is the orbit Gx containing x. Further, σ and τ have the same image if and only if $\sigma x = \tau x$, if and only if $\tau^{-1}\sigma x = x$, if and only if $\tau^{-1}\sigma \in G_x$, if and only if $\sigma G_x = \tau G_x$. Therefore the map $G \to X$ induces an isomorphism

$$G/G_x \xrightarrow{\sim} Gx, \qquad \sigma G_x \mapsto \sigma x$$

of G-sets. Thus every orbit is G-isomorphic to a left coset space for a subgroup. Counting the cardinalities of both sets one finds that

$$\#Gx = (G : G_x).$$

Transitive G-sets. Let G be a group. A G-set X is called *transitive* if there is exactly one orbit of X under G; or, equivalently, if $X \neq \emptyset$ and for any two elements $x, y \in X$ there exists $\sigma \in G$ with $\sigma x = y$. For example, if $X = G/H$ for some subgroup $H \subset G$, then G/H is a transitive G-set. The following result expresses that this is the *only* example, up to isomorphism.

Proposition. *For every transitive G-set X there is a subgroup H of G with $X \cong_G G/H$, and X determines H uniquely up to conjugacy.*

Proof. Let $x \in X$. Then we have $X = Gx$, so by what we proved above we have $X = Gx \cong_G G/G_x$, which proves the first assertion, with $H = G_x$. Replacing x by τx one sees that one also has $X = G\tau x \cong_G G/G_{\tau x} = G/\tau G_x \tau^{-1}$, so G_x may be replaced by any conjugate. Conversely, suppose that K is a subgroup of G with $X \cong_G G/K$. Since K is the stabilizer of an element of G/K, it is also the stabilizer of an element of X; the latter element is of the form τx for some $\tau \in G$, so we have $K = G_{\tau x} = \tau G_x \tau^{-1}$. This proves the proposition.

General G-sets. Let G be a group. The proposition just proved shows that if we "know" all subgroups of G, then we "know" all transitive G-sets. We now pass to general G-sets. An arbitrary G-set X can be written as the disjoint union of its orbits, each of which is a transitive G-set. Thus, if $R \subset X$ is a set of representatives for the orbits, then there is a G-isomorphism

$$X \cong_G \coprod_{x \in R} G/G_x,$$

where $\coprod$ denotes disjoint union. Counting the cardinalities of both sets one finds the important formula

$$\#X = \sum_{x \in R} (G : G_x).$$

Example. Let G be a group, acting on itself by conjugation. The orbits are then called the *conjugacy classes* of G. The stabilizer of $x \in G$ is equal to the *centralizer* $C_G(x) = \{g \in G : gx = xg\}$ of x in G. The cardinality of the conjugacy class containing x equals $\big(G : C_G(x)\big)$, and if $R \subset G$ contains exactly one element from each conjugacy class, then we have

$$\#G = \sum_{x \in R} \big(G : C_G(x)\big).$$

2. Exercises

2.1. Let G be a group and X a set.

(a) Let $G \times X \to X$ be a map, written as $(\sigma, x) \mapsto x\sigma$. Prove that this map is a right action if and only if the map $G \times X \to X$ sending $(\sigma, x) \mapsto x\sigma^{-1}$ is a left action.

(b) Let $G \times X \to X$ be a left action. Prove that it is a right action if and only if the induced action of $[G, G]$ on X is trivial.

2.2. Prove: (a) the trivial action of a group G on a set X is free if and only if $G = \{1\}$ or X is empty; (b) any free action of a group on a non-empty set is faithful; (c) an action of a group G on a set X is faithful if and only if the corresponding map $G \to \operatorname{Sym} X$ is injective; (d) if X is a trivial, free, or faithful G-set, then for any subgroup H of G it is also trivial, free, or faithful (respectively) when considered as an H-set.

2.3. Let X be the set $\{1, 2, 3, 4\}$, which is naturally acted upon by S_4. Let Y be the set of all splittings of X into two subsets of size 2, i. e.

$$Y = \big\{\{A, B\} : A, B \subset X, A \cap B = \emptyset, \#A = \#B = 2\big\}.$$

(a) Determine $\#Y$.

(b) Show that from the action of S_4 on X one obtains in a natural way an action of S_4 on Y, and that the corresponding map $S_4 \to \operatorname{Sym} Y$ is surjective. What is the kernel of this map?

2.4. Let G be a group, and let X and Y be G-sets. Write Y^X for the set of all maps $X \to Y$. Show that Y^X is a G-set, the action being defined by $(\sigma f)(x) = \sigma\big(f(\sigma^{-1}x)\big)$, for $\sigma \in G$, $f \in Y^X$, $x \in X$. Prove also that $\operatorname{Fix}_G(Y^X)$ is the set of G-maps $X \to Y$.

2.5. Let G be a group, $N \subset G$ a normal subgroup, and X a G-set.

(a) Prove that there are unique actions of G on X^N and on $N\backslash X$ for which the inclusion map $X^N \subset X$ and the natural map $X \to N\backslash X$ are G-maps. Show also that these actions are induced by actions of G/N on X^N and on $N\backslash X$, via the canonical map $G \to G/N$.

(b) Suppose that $H \subset G$ is a subgroup, and let X be the G-set G/H. Suppose that there is a G-action on X^H for which the inclusion $X^H \subset X$ is a G-map, or that there is a G-action on $H\backslash X$ for which the natural map $X \to H\backslash X$ is a G-map. Prove that H is normal in G.

2.6. Let G be a group acting on a set X. Prove that the number of G-subsets of X equals $2^{\#(G\backslash X)}$.

2.7. Suppose that the quaternion group Q of order 8 acts faithfully on a set X. Prove: $X \geq 8$. Which other groups of order 8 have this property?

2.8. Let G be a group, and let X be a G-set. We call X *primitive* if (i) X is transitive, (ii) $\#X > 1$, and (iii) if Y is a G-set with $\#Y > 1$, and $f: X \to Y$ is a surjective G-map, then f is bijective.

Prove that X is primitive if and only if there is a maximal subgroup H of G with $X \cong_G G/H$; here a subgroup H of a group G is called *maximal* if there are precisely two different subgroups K of G with $H \subset K \subset G$ (namely, $K = H$ and $K = G$).

2.9. Let G be a group, let H be a subgroup of G, and let N be the kernel of the group homomorphism $G \to \mathrm{Sym}(G/H)$ corresponding to the natural action of G on G/H. Prove that N is a normal subgroup of G that is contained in H, and that every normal subgroup of G contained in H is contained in N.

2.10. Let G be a group, and let H be a subgroup of finite index n in G. Prove that G has a normal subgroup N contained in H for which $(H : N)$ divides $(n-1)!$. What does this come down to in the case $n = 2$?

2.11. Let G be a finite group of order greater than 1, and let p be the smallest prime number dividing the order of G.

(a) Prove: any subgroup of G of index p is normal in G and contains $[G, G]$.

(b) Prove: any normal subgroup of order p is contained in $Z(G)$.

2.12. Decide whether the following statement is true, by giving a proof or a counterexample. Let G be a group, H a subgroup of finite index n, and let m be a positive integer. Suppose that for all $h \in H$ one has $h^m = e$. Then for all $g \in G$ one has $g^{nm} = e$.

2.13. (a) Determine all finite groups, up to isomorphism, that have at most 3 conjugacy classes.

(b) Let k be a positive integer. Prove that, up to isomorphism, there are only finitely many finite groups that have exactly k conjugacy classes.

2.14 (Burnside's lemma). Let G be a finite group acting on a finite set X. Prove:

$$\#(G\backslash X) = \frac{1}{\#G} \cdot \sum_{\sigma \in G} \#\{x \in X : \sigma x = x\}.$$

2.15. If a group G acts on a finite set X, and an element σ of G induces a permutation of X that is a product of t pairwise disjoint cycles of lengths n_1, n_2, $\ldots$, n_t (also counting cycles of length 1), then we say that σ has *type* $(n_1, n_2, \ldots, n_t)$ in its action on X.

Let $D \subset \mathbf{R}^3$ be a regular dodecahedron, and let G be the group of rotations σ of 3-space with $\sigma D = D$. (No reflections are allowed; so if 0 is the center of gravity of D, then each $\sigma \in G$ is given by an orthogonal 3×3-matrix with determinant $+1$.) Denote by F the set of 2-dimensional faces of D; so $\#F = 12$.

(a) Argue geometrically that G is a finite group of order 60, and that G, in its action on F, has

1 element of type $(1,1,1,1,1,1,1,1,1,1,1,1)$,
15 elements of type $(2,2,2,2,2,2)$,
20 elements of type $(3,3,3,3)$, and
24 elements of type $(1,1,5,5)$.

(b) (*Counting colored dodecahedra.*) Let F be as above, and let C be a finite set (of "colors"). Let G act trivially on C, and let G act on the set C^F of functions $F \to C$ as in Exercise 2.4. One can think of an element of C^F as a way of coloring the faces of D with the colors from C; then two elements of C^F are in the same G-orbit if and only if the corresponding colored dodecahedra are the same up to rotation. Thus $\#(G\backslash C^F)$ is the number of "essentially" different colorings. Prove that this number is given by

$$\frac{n^{12}+15n^6+44n^4}{60}.$$

2.16. Prove that the group G from Exercise 2.15 is isomorphic to the alternating group A_5.

2.17. Let G be a group, and let H be a subgroup of G. The *normalizer* $N_G(H)$ of H in G is defined by $N_G(H) = \{g \in G : gH = Hg\}$, and the *centralizer* $C_G(H)$ of H by $C_G(H) = \{g \in G : \text{for all } x \in H \text{ one has } gx = xg\} = \bigcap_{x\in H} C_G(x)$.

(a) Let G act by conjugation on the set of subgroups of G. Prove that $N_G(H)$ is the stabilizer of H under this action.

(b) Prove: H and $C_G(H)$ are normal subgroups of $N_G(H)$, and $N_G(H)/C_G(H)$ is isomorphic to a subgroup of $\operatorname{Aut} H$.

2.18 (the Burnside ring). Let G be a group, and let X, Y be G-sets. The *disjoint union* $X \amalg Y$ is the G-set whose underlying set is the disjoint union of X and Y, the G-action being the unique one for which the two natural inclusions $X \to X \amalg Y$ and $Y \to X \amalg Y$ are G-maps. The *product* $X \times Y$ is the G-set whose underlying set is the cartesian product of X and Y, the G-action being the unique one for which the two natural projections $X \times Y \to X$ and $X \times Y \to Y$ are G-maps. Now let G be finite. Write $B'(G)$ for the set of G-isomorphism classes of finite G-sets, the isomorphism class of X being denoted by $[X]$. Define an addition and a multiplication on $B'(G)$ by

$$[X]+[Y]=[X \amalg Y], \qquad [X]\cdot[Y]=[X\times Y].$$

(a) Prove that $B'(G)$ satisfies all usual axioms for a commutative ring, with the exception of the existence of additive inverses. What is the zero element of $B'(G)$? What is the unit element?

(b) Construct a ring $B(G)$ by formally adjoining negatives, imitating the construction of the field of fractions. This ring is called the *Burnside ring* of G. To which known ring is $B(G)$ isomorphic if $G = \{1\}$?

(c) Prove that the additive group of $B(G)$ is isomorphic to $\mathbf{Z}^k$, where k is the number of conjugacy classes of subgroups of G.

2.19. Let G be a finite group, and keep the notation from Exercise 2.18.

(a) Prove: for every subgroup $H \subset G$ there is a ring homomorphism $B(G) \to \mathbf{Z}$ that maps the class $[X]$ of any finite G-set X to $\#\operatorname{Fix}_H X$.

(b) Prove: if two finite G-sets X and Y are such that for every subgroup $H \subset G$ one has $\#\operatorname{Fix}_H X = \#\operatorname{Fix}_H Y$, then one has $X \cong_G Y$.

(c) Let k be as in Exercise 2.18(c). Prove that there is an injective ring homomorphism $B(G) \to \mathbf{Z}^k$, the ring operations in $\mathbf{Z}^k$ being componentwise. What is the image of this map if G is cyclic of prime order?

2.20. Let G and J be groups, and let Y be a J-set. The *wreath product* $G \wr J$ of G and J (relative to Y) is defined as follows. Let G^Y be the set of functions $f\colon Y \to G$, made into a group by $(f_1 f_2)(y) = f_1(y) f_2(y)$, for $f_1, f_2 \in G^Y$, $y \in Y$. For $\sigma \in J$, $f \in G^Y$, let ${}^{\sigma}f \in G^Y$ map y to $f(\sigma^{-1}y)$ (cf. Exercise 2.4). Now $G \wr J$ is defined to be the semidirect product $G^Y \rtimes J$ (see Exercise 1.6).

(a) Express $\#(G \wr J)$ is terms of $\#G$, $\#J$, and $\#Y$.

(b) Suppose that $\#G = \#J = \#Y = 2$, the action of J on Y being non-trivial. Prove: $G \wr J \cong D_4$.

(c) Let X be a G-set. Prove that $X \times Y$ becomes a $G \wr J$-set through

$$(f, \sigma)(x, y) = (f(y)x, \sigma y),$$

for $f \in G^Y$, $\sigma \in J$, $x \in X$, $y \in Y$.

(d) Prove: $X \times Y$ is a transitive $G \wr J$-set if and only if X is a transitive G-set and Y is a transitive J-set; and if $Y \neq \emptyset$, then $X \times Y$ is a faithful $G \wr J$-set if and only if X is a faithful G-set and Y is a faithful J-set.

2.21. Let G be a group acting transitively on a set X.

(a) Let $H \subset G$ be the stabilizer of some $x \in X$. Prove that the group $\operatorname{Aut}_G X$ of G-automorphisms of X is isomorphic to $N_G(H)/H$, where $N_G(H)$ is the normalizer of H in G (see Exercise 2.17).

(b) Let G act on itself by left multiplication. Prove: $\operatorname{Aut}_G G \cong G$.

(c) Let n be a positive integer, and let the G-set Y be the disjoint union of n copies of X. Prove that $\operatorname{Aut}_G Y$ is isomorphic to the wreath product $(\operatorname{Aut}_G X) \wr S_n$ relative to the S_n-set $\{1, 2, \ldots, n\}$ (see Exercise 2.20 for wreath products).

2.22. Let G be a finite group. Prove the identity

$$\sum_Y \frac{t^{\# Y}}{\#\operatorname{Aut}_G Y} = \exp\Big(\sum_X \frac{t^{\# X}}{\#\operatorname{Aut}_G X}\Big)$$

in the ring $\mathbf{Q}[[t]]$ of formal power series in t with rational coefficients; here Y ranges over all finite G-sets, up to isomorphism, X ranges over all transitive finite G-sets, up to isomorphism, and for $f \in t\mathbf{Q}[[t]]$ one defines $\exp f = \sum_{n=0}^{\infty} f^n/n! \in 1 + t\mathbf{Q}[[t]]$.

2.23. (a) Let G be a group acting transitively on a set X with $\#X > 1$, and let $H \subset G$ be the stabilizer of some $x \in X$. Let the G-set Y be the disjoint union of two copies of X. Prove that Y has an H-automorphism that is not a G-automorphism.

(b) Let G be a group, and let $H \subset G$ be a subgroup with $H \neq G$. Construct a group K and two group homomorphisms $G \to K$ that agree on H but not on all of G.

2.24. Let G be a group, and let X be a finite G-set. As in Exercise 2.4, let $X^{\mathbf{Z}}$ be the G-set of all maps $\mathbf{Z} \to X$, with G acting trivially on $\mathbf{Z}$. We call an element $f \in X^{\mathbf{Z}}$ a *cycle* if there exist $\sigma \in G$ and $x \in X$ such that for all $i \in \mathbf{Z}$ one has $f(i) = \sigma^i x$.

(a) Prove that the set C of cycles is a G-subset of $X^{\mathbf{Z}}$.

(b) Prove: $\#(G\backslash C) = \#X$.

2.25. Let G be a group, and let $H \subset G$ be a subgroup. For $x, y \in G$ we write $x \sim_H y$ to mean that there exists $z \in H$ such that for all $i \in \mathbf{Z}$ one has $x^i H = z y^i H$.

(a) Prove that $\sim_H$ is an equivalence relation on G, that H is one of the equivalence classes, and that for $x, y \in G$ one has $x \sim_H y$ if and only if $x^{-1} \sim_H y^{-1}$.

(b) Prove that H is normal in G if and only if for all $x, y \in G$ one has: $x \sim_H y \Leftrightarrow xH = yH$.

(c) Prove that every equivalence class of $\sim_H$ has the same cardinality as H.

2.26. Let $f\colon G_1 \to G_2$ be a group homomorphism.

(a) The *graph* Γ_f of f is defined by $\Gamma_f = \{(x, f(x)) : x \in G_1\} \subset G_1 \times G_2$. Prove that this is a subgroup of $G_1 \times G_2$.

(b) Suppose that G_1 and G_2 are finite, and let n be a positive integer. Prove that $\#\{(x,y) \in G_1 \times G_2 : f(x^n) = y^n\} \equiv 0 \bmod \#G_1$.

3. Sylow theorems

Let p be a prime number. A *p-group* is a finite group of which the order is a power of p; the zeroth power $p^0 = 1$ is allowed, so the trivial group $\{1\}$ is a p-group for all primes p. Other example: the Klein four group V_4 is a 2-group.

A p-subgroup of a group is a subgroup that is a p-group. Sylow theory studies p-subgroups of finite groups. Examples: V_4 and D_4 are 2-subgroups of S_4; for any prime number p and any integer $n \geq 0$ the set of upper triangular $n \times n$-matrices over $\mathbf{F}_p$ with 1's on the diagonal is a p-subgroup of $\mathrm{Gl}(n, \mathbf{F}_p)$.

Theorem 3.1. *Let p be a prime number, G a finite group, H a p-subgroup, and m a non-negative integer. Suppose that $(G : H)$ is of the form np^m for some positive integer n. Then G has a subgroup J with $H \subset J \subset G$ and $\#J = p^m \cdot \#H$, and the number of such subgroups J is $1 \bmod p$.*

Proof. The second statement implies the first, since any number that is $1 \bmod p$ is non-zero. So it suffices to prove the second statement. For this purpose, let

$$X = \{S \subset G : \#S = p^m \cdot \#H,\ SH = S\},$$

where $SH = \{sh : s \in S, h \in H\}$. We count X in two ways. The first is by a direct combinatorial argument. The condition $SH = S$ is equivalent to S being a union of left cosets aH; and $\#S = p^m \cdot \#H$ then means that it is a union of p^m different such cosets. Thus $\#X$ is the same as the number of subsets of G/H of cardinality p^m, and by $\#(G/H) = np^m$ we find

$$\#X = \binom{np^m}{p^m}.$$

The second way of counting X makes use of the action of G on X defined by $gS = \{gs : s \in S\}$; this is indeed a well-defined action, since $\#gS = \#S = p^m$ and $gSH = gS$. As we saw in the previous section, $\#X$ equals the sum of the cardinalities of the orbits of this action. Fix $S \in X$ for the moment. The orbit containing S has cardinality $(G : G_S)$, where G_S is the stabilizer of S. Since S is closed under left multiplication by elements of G_S, it is the union of a certain number of right cosets $G_S a$ of G_S. Because $\#S = p^m \cdot \#H$ is a power of p, that number of cosets is a power of p as well; say it is $p^{i(S)}$, with $i(S) \geq 0$. Then we have $\#G_S = \#S/p^{i(S)} = p^m \cdot \#H/p^{i(S)}$, and the cardinality of the orbit is given by

$$(G : G_S) = \#G/\#G_S = np^{i(S)}.$$

Now let S range over a set R of representatives for the orbits. Adding up the numbers $(G : G_S)$ we obtain $\#X$, so

$$\binom{np^m}{p^m} = \sum_{S \in R} np^{i(S)}.$$

If we take this modulo np, then on the right only those S remain for which $i(S) = 0$; these come in orbits of cardinality n, so we obtain

$$\binom{np^m}{p^m} \equiv \#\{S \in X : i(S) = 0\} \bmod np.$$

Let again $S \in X$. By definition of $i(S)$, we have $i(S) = 0$ if and only if S consists of a single right coset of G_S, and this occurs if and only if S is a right coset Ka of some subgroup $K \subset G$ of order $p^m \cdot \#H$ (then $G_S = K$). But a subset of a group is a right coset of some subgroup if and only if it is a left coset of some subgroup (proof: $Ka = a \cdot a^{-1}Ka$). Hence for $S \in X$ we have:

$$i(S) = 0 \Leftrightarrow S \text{ is a left coset } aJ \text{ of some subgroup } J \text{ of } G \text{ of order } p^m \cdot \#H.$$

If $J \subset G$ is a subgroup of G of order $p^m \cdot \#H$, then a left coset aJ belongs to X if and only if $J \supset H$, as one readily checks. Since each J gives n cosets we can now conclude:

$$(*) \quad \binom{np^m}{p^m} \equiv n \cdot \#\{J : J \text{ is a subgroup of order } p^m \cdot \#H \text{ of } G \text{ containing } H\} \bmod np.$$

This holds whenever p, G, H, m, n satisfy the hypotheses of the theorem. In particular, when we fix p, m, and n, it holds if G is cyclic of order np^m and $H = \{1\}$; in that case, there is exactly one subgroup J of order p^m, so $(*)$ tells us

$$\binom{np^m}{p^m} \equiv n \bmod np.$$

Now that we know what $\binom{np^m}{p^m}$ modulo np is, we go back to $(*)$ for the case of general G and H, and we find that

$$(*) \qquad n \equiv n \cdot \#\{J : J \text{ is a subgroup of order } p^m \cdot \#H \text{ of } G \text{ containing } H\} \bmod np.$$

Dividing by n we obtain the second assertion of the theorem. This proves 3.1.

The idea of the proof just given is due to H. Wielandt; the trick with the cyclic group was invented by J. McLaughlin.

Corollary 3.2. *If G is a finite group and q is a prime power dividing $\#G$, then G has a subgroup of order q.*

Proof. Apply Theorem 3.1 to $H = \{1\}$ and $p^m = q$. This proves 3.2.

The condition that q be a prime power cannot be omitted in 3.2: see Exercises 3.1 and 3.21.

Let p be a prime number, and let G be a finite group. A *Sylow p-subgroup* (or *p-Sylow subgroup*) of G is a subgroup of G whose order equals the largest power of p dividing the order of G (Ludvig Sylow, Norwegian mathematician, 1832–1918). It follows from 3.2 that every finite group has a Sylow p-subgroup for each prime number p. The following result gives more information about Sylow subgroups.

Theorem 3.3. *Let p be a prime number and let G be a finite group. Then we have:*
(a) *the number of Sylow p-subgroups of G is a divisor of $\#G$ that is $1 \bmod p$;*
(b) *each p-subgroup of G is contained in some Sylow p-subgroup of G;*
(c) *any two Sylow p-subgroups of G are conjugate in G.*

Proof. Assertion (b) follows from Theorem 3.1, with H equal to the given p-subgroup and m such that $p^m \cdot \#H$ equals the highest power of p dividing $\#G$. For (c), let S_1 and S_2 be two Sylow p-subgroups of G, and consider the natural action of S_1 on G/S_2. Each orbit has cardinality equal to the index of some subgroup of S_1, which is a power of p; that power of p is either divisible by p or equal to 1, and the latter occurs only for the fixed points. Thus, adding up the lengths of all orbits, we obtain

$$\#(G/S_2) \equiv \# \operatorname{Fix}_{S_1}(G/S_2) \bmod p.$$

Since $\#S_2$ is the highest power of p dividing $\#G$, the number on the left is non-zero modulo p. Hence the number on the right is non-zero, i. e., the group S_1 fixes some element aS_2 of G/S_2 and is therefore contained in the stabilizer of aS_2 in G, which is aS_2a^{-1}. But S_1 and S_2 have the same order, so we actually have $S_1 = aS_2a^{-1}$. This proves (c).

Let G act by conjugation on the set of all of its subgroups. From (c) we see that the set of Sylow p-subgroups is a single orbit under this action. Therefore the number of Sylow p-subgroups equals the cardinality of that orbit, which by Section 2 is a divisor of $\#G$ (it is actually equal to $\big(G : N_G(S_1)\big)$, the normalizer $N_G(S_1)$ being defined as in Exercise 2.17). This proves the first part of (a); the second part of (a) follows from Theorem 3.1, with $H = \{1\}$ and p^m equal to the highest power of p dividing $\#G$. This proves 3.3.

Theorems stating that finite groups have "many" p-subgroups are useful because p-groups are usually much easier to handle than general finite groups. See Exercises 3.10, 3.12, and 3.13 for a few strong properties of p-groups.

Example: groups of order 42. We illustrate the usefulness of Sylow subgroups in determining the structure of finite groups. Let G be a group of order $42 = 2 \cdot 3 \cdot 7$, and for each $p \in \{2, 3, 7\}$ let $S(p)$ be a Sylow p-subgroup of G; it is cyclic of order p. By Theorem 3.3, the number of subgroups of G conjugate to $S(7)$ is a divisor of 42 that is 1 mod 7; the only such divisor is 1, so $S(7)$ is a *normal* subgroup of G. Therefore $S(3)S(7)$ is a subgroup of order 21 of G, and applying Exercise 1.7(a)—with $S(3)S(7)$, $S(3)$, $S(7)$ in the roles of G, H, N respectively—one deduces that $S(3)S(7)$ is the semidirect product of $S(3)$ and $S(7)$ with respect to some group homomorphism $S(3) \to \operatorname{Aut} S(7)$; here $\operatorname{Aut} S(7) \cong \mathbf{F}_7^*$ is cyclic of order 6. If that group homomorphism is trivial then $S(3)S(7)$ is cyclic of order 21; if it is non-trivial then its image is the unique subgroup of $\operatorname{Aut} S(7)$ of order 3, and $S(3)S(7)$ is the unique non-abelian group of order 21 (up to isomorphism; cf. Exercise 3.5(b)). In both cases, $S(3)S(7)$ is normal in G because it is of index 2. Again by Exercise 1.7(a), now with $S(2)$ and $S(3)S(7)$ in the roles of H and N, the entire group G is the semidirect product of

$S(3)S(7)$ and $S(2)$ with respect to some group homomorphism $S(2) \to \mathrm{Aut}\big(S(3)S(7)\big)$. If that group homomorphism is trivial then G is the product of $S(3)S(7)$ and a cyclic group of order 2; this gives two possibilities for G, up to isomorphism. Next suppose that the map $S(2) \to \mathrm{Aut}\big(S(3)S(7)\big)$ is non-trivial. Its image is a subgroup of order 2 of $\mathrm{Aut}\big(S(3)S(7)\big)$; if $S(3)S(7)$ is abelian, then the group $\mathrm{Aut}\big(S(3)S(7)\big) \cong (\mathbf{Z}/21\mathbf{Z})^* \cong \mathbf{F}_3^* \times \mathbf{F}_7^*$ has three subgroups of order 2, which gives three possibilities for G. If $S(3)S(7)$ is non-abelian, then one checks that $\mathrm{Aut}\big(S(3)S(7)\big)$ has, up to conjugacy, only one subgroup of order 2; and from this one deduces that there is only one possibility left for G, up to isomorphism. Modulo this verification, and the verification that all groups obtained are pairwise non-isomorphic, we conclude that there are exactly six groups of order 42, up to isomorphism.

Often one can simplify the work by invoking additional theorems. For example, 42 is squarefree, and any group of squarefree order is the semidirect product of two cyclic groups (Exercise 3.27).

3. Exercises

3.1. Prove that the alternating group A_4 of order 12 does not have a subgroup of order 6.

3.2. Let G be a finite group, and let for each prime number p a Sylow p-subgroup $S(p)$ of G be chosen. Prove that G is generated by the subgroups $S(p)$, as p ranges over all prime numbers.

3.3. Let G be a finite group, and let $S \subset G$ be a Sylow p-subgroup for some prime number p. Prove: S is normal in G if and only if S is characteristic in G.

3.4. (a) Prove that, up to isomorphism, there are precisely 19 groups of order smaller than 12. Which are they?

(b) Let G be a group of order 12. Prove that G has a normal Sylow p-subgroup for some $p \in \{2, 3\}$.

(c) How many pairwise non-isomorphic groups of order 12 are there? Prove the correctness of your answer.

3.5. Let p and q be prime numbers with $p > q$.

(a) Prove: if G is a group of order pq, then G has a normal Sylow p-subgroup.

(b) Prove: if q does not divide $p - 1$, then every group of order pq is cyclic; and if q does divide $p - 1$, then there is a unique non-cyclic group of order pq, up to isomorphism, and it is non-abelian.

3.6. Prove that any group of order 1001 is cyclic.

3.7. A group is called *solvable* if it has a normal tower $(u_i)_{i=0}^r$ (see Exercise 1.3) for which every group u_i/u_{i-1} is abelian.

(a) Prove: if G is a solvable group, then every subgroup H is solvable, and G/N is solvable for every normal subgroup N of G.

(b) Prove: every solvable group G has a normal tower $(v_i)_{i=0}^t$ for which every v_i is characteristic in G and every group v_i/v_{i-1} is abelian.

3.8. (a) Prove that every group of order smaller than 60 is solvable.

(b) Prove that the group A_5 is a group of order 60 that is not solvable.

3.9. Let G be a group of order 60 that is not solvable. The purpose of this exercise is to show that G is isomorphic to A_5.

(a) Prove that G is simple (see Exercise 1.4) and that G has precisely six Sylow 5-subgroups.

(b) Using the action of G on the set of its Sylow 5-subgroups, show that there is an embedding of G as a subgroup in A_6.

(c) Using the action of A_6 on A_6/G, show that there is group homomorphism $A_6 \to S_6$ that induces an isomorphism $G \to A_5$; here A_5 is identified with the subgroup of S_6 consisting of all even permutations of $\{1, 2, 3, 4, 5, 6\}$ that fix 6.

3.10. Let p be a prime number, and let G be a p-group.

(a) Prove: every maximal subgroup of G (as in Exercise 2.8) has index p and is normal in G. (*Hint*: use Exercise 2.11(a).)

(b) Prove that G is solvable.

3.11. Let p be a prime number, G a p-group, and X a finite G-set. Prove: $\#X^G \equiv \#X \bmod p$; here $X^G = \operatorname{Fix}_G X$.

3.12. Let p be a prime number.

(a) Prove: if G is a p-group of order greater than 1, then $Z(G) \neq \{1\}$.

(b) Prove: if G is a p-group, and $N \subset G$ is a normal subgroup of order greater than 1, than $N \cap Z(G) \neq \{1\}$.

3.13. A group G is called *supersolvable* is it has a normal tower $(u_i)_{i=0}^r$ for which every u_i is normal in G and every group u_i/u_{i-1} is cyclic.

(a) Give an example of a finite group that is solvable but not supersolvable.

(b) Prove: if p is a prime number, then every p-group is supersolvable.

3.14. (a) Prove that for every prime number p there are exactly two groups of order p^2, up to isomorphism, and that they are both abelian.

(b) Prove that for every prime number p there are exactly five groups of order p^3, up to isomorphism.

3.15. Let p be a prime number, G a finite group, and S a Sylow p-subgroup of G.

(a) Prove: if N a normal subgroup of G, then $S \cap N$ is a Sylow p-subgroup of N, and the image of S in G/N is a Sylow p-subgroup of G/N.

(b) Prove: if H is a subgroup of G, then there is a conjugate T of S in G such that $T \cap H$ is a Sylow p-subgroup of H. Can one always take $T = S$? Give a proof or a counterexample.

3.16. Let p, G, and S be as in the previous exercise, and let N be a normal subgroup of G.

(a) Prove: if $S \cap N$ is normal in N, then it is also normal in G.

(b) Prove: $N_G\big(N_G(S)\big) = N_G(S)$, the notation being as in Exercise 2.17.

3.17 (Frattini argument). Let p be a prime number, G a finite group, N a normal subgroup of G, and P a Sylow p-subgroup of N. Prove: $N \cdot N_G(P) = G$.

3.18. Let k be a finite field, and denote by p its characteristic. Let $G = k^+ \rtimes k^*$ be the semidirect product of the additive group k^+ and the multiplicative group k^* of k, the action of the latter on the former being multiplication in k.

(a) Prove that G is isomorphic to the subgroup $\{\left(\begin{smallmatrix} a & b \\ 0 & 1 \end{smallmatrix}\right) : a \in k^*,\, b \in k\}$ of $\mathrm{Gl}(2, k)$.

(b) Prove that G has a normal Sylow p-subgroup.

(c) Let l be a prime number different from p. Prove that G has a cyclic Sylow l-subgroup. For which l is the Sylow l-subgroup normal in G?

3.19. Supply the details of the classification of groups of order 42 given in the text. Where does the cyclic group of order 42 appear in this classification? And the dihedral group D_{21}? And the group $\mathbf{F}_7 \rtimes \mathbf{F}_7^*$ from Exercise 3.18?

3.20. Let k and G be as in Exercise 3.18, and let n be a positive integer. Prove that G has a subgroup of order n if and only if there are positive integers n_1 and n_2 satisfying

$$n = n_1 n_2, \qquad n_1 \text{ divides } \#k, \qquad n_2 \text{ divides } \#k^*, \qquad n_1 \equiv 1 \bmod n_2.$$

3.21. Suppose that n is a positive integer with the property that every finite group of order divisible by n has a subgroup of order n. Prove that n is a prime power. (*Hint*: use Exercise 3.20.)

3.22 (Sylow subgroups of linear groups). Let n be a non-negative integer, k a finite field, and p and q the characteristic and the cardinality of k, respectively.

(a) Prove: $\#\,\mathrm{Gl}(n, k) = \prod_{i=0}^{n-1} (q^n - q^i)$.

(b) Prove that the upper triangular matrices in $\mathrm{Gl}(n, k)$ with 1's on the diagonal form a Sylow p-subgroup of $\mathrm{Gl}(n, k)$.

3.23 (Sylow subgroups of S_n). Let p be a prime number, and let n be a non-negative integer. Write n in base p, i. e., write $n = \sum_{i \ge 0} c_i p^i$ with $c_i \in \{0, 1, \ldots, p-1\}$ for all i and $c_i = 0$ for all but finitely many i.

(a) Prove that the exponent of the highest power of p dividing $n!$ is $\left(n-\sum_i c_i\right)/(p-1)$.

(b) Let k be a positive integer, and denote by C_p a cyclic group of order p. Denote by $W(p,k)$ the iterated wreath product $(\cdots((C_p \wr C_p) \wr C_p)\cdots) \wr C_p$; here the sign $\wr$ appears $k-1$ times, C_p appears k times, and each wreath product (see Exercise 2.20) is taken relative to the C_p-set C_p, the action being left multiplication. Give a formula for the order of $W(p,k)$, and prove that $W(p,k)$ is isomorphic to a Sylow p-subgroup of the symmetric group S_{p^k}.

(c) Describe a Sylow p-subgroup of the symmetric group S_n in terms of the groups $W(p,i)$ and the numbers c_i.

3.24. Let G be a finite group. Let G act on itself by left multiplication, and let $G \to \operatorname{Sym} G$ be the corresponding group homomorphism.

(a) Prove: the image of G in $\operatorname{Sym} G$ contains an odd permutation if and only if any Sylow 2-subgroup of G is cyclic and non-trivial.

(b) Prove: if G has a cyclic Sylow 2-subgroup, then $N = \{x \in G : x \text{ has odd order}\}$ is a characteristic subgroup of odd order of G, and G/N is cyclic of 2-power order.

3.25 (the Verlagerung). Let G be a group, and let H be a subgroup of finite index n of G. We write $\bar{H}$ for the abelian group $H/[H,H]$, and ρ for the natural map $H \to \bar{H}$. The *Verlagerung* (or *transfer*) $\operatorname{Ver}_{G\to H}\colon G \to \bar{H}$ is defined as follows. Write $G = a_1H \cup a_2H \cup \cdots \cup a_nH$, and for each $g \in G$ and $i \in \{1,2,\ldots,n\}$ let $j = j(g,i)$ be such that $ga_i \in a_jH$, say $ga_i = a_jh_{g,i}$, with $h_{g,i} \in H$; then $\operatorname{Ver}_{G\to H}(g) = \prod_{i=1}^{n} \rho(h_{g,i}) \in \bar{H}$.

(a) Prove: this definition is independent of the choice of the coset representatives a_i, and $\operatorname{Ver}_{G\to H}$ is a group homomorphism.

(b) Let $g \in G$ have type $\big(n(1),\ldots,n(t)\big)$ in its action on G/H, as in Exercise 2.15; for each $j \in \{1,2,\ldots,t\}$, let $a_{i(j)}H$ belong to the jth cycle, which has length $n(j)$. Prove that one has

$$\operatorname{Ver}_{G\to H}(g) = \prod_{i=1}^{t} \rho\big(a_{i(j)}^{-1} g^{n(j)} a_{i(j)}\big).$$

3.26. Let G be a finite group of order greater than 1, and let p be the smallest prime number dividing $\#G$. Let S be a Sylow p-subgroup of G, and suppose that S is *cyclic*.

(a) Let $H \subset S$ be a subgroup. Prove that the normalizer $N_G(H)$ of H in G is equal to the centralizer $C_G(H)$ of H in G.

(b) Suppose that x, y are elements of S that are conjugate in G. Prove: $x = y$.

(c) Let $\operatorname{Ver}_{G\to S}\colon G \to S/[S,S] = S$ be as in Exercise 3.25. Prove that the restriction of $\operatorname{Ver}_{G\to S}$ to S is an automorphism of S, and that $N = \{x \in G : x \text{ has order not divisible by } p\}$ is a characteristic subgroup of G with $G/N \cong S$.

3.27. Let G be a finite group of squarefree order.

(a) Prove that G is solvable. (*Hint*: use Exercise 3.26(c).)
(b) Prove that $G'' = \{1\}$, the notation being as in Exercise 1.20.
(c) Prove that G is the semidirect product of two cyclic groups.

3.28. Let n be the product of all prime numbers p with $50 < p < 100$. Prove that every group of order n is cyclic.

3.29. Let n be a positive integer, and write $\varphi(n) = \#(\mathbf{Z}/n\mathbf{Z})^*$. Prove: every group of order n is cyclic if and only if n and $\varphi(n)$ are relatively prime.

4. Exact sequences

In this section we restrict to *abelian* groups (but see Exercise 4.1), and we write them additively. A homomorphism $f\colon A \to B$ of abelian groups will often be written as $A \xrightarrow{f} B$, or simply as $A \to B$ if the map is clear from the context, or no notation is required for it. We call a sequence

$$A \xrightarrow{f} B \xrightarrow{g} C$$

of homomorphisms *exact* if $\operatorname{im} f = \ker g$. Denoting the zero group by 0, we have:

(i) $0 \to B \xrightarrow{g} C$ is an exact sequence if and only if g is injective;
(ii) $A \xrightarrow{f} B \to 0$ is exact if and only if f is surjective;
(iii) $0 \to A \xrightarrow{f} B \to 0$ is exact if and only if f is an isomorphism;
(iv) $0 \to B \to 0$ is exact if and only if $B = 0$.

In general, a sequence of homomorphisms $A_1 \to A_2 \to A_3 \to \ldots \to A_n$ is called *exact at* A_i (where $1 < i < n$) if $A_{i-1} \to A_i \to A_{i+1}$ is exact. The entire sequence is exact if it is exact at each A_i for $1 < i < n$. We use the same terminology for infinite sequences $A_1 \to A_2 \to A_3 \to \ldots$, $\ldots \to A_1 \to A_2 \to A_3 \to \ldots$, etc. An exact sequence of the form $0 \to A \to B \to C \to 0$ is known as a *short exact sequence*. In this case, $A \to B$ is injective, so that we can view A as a (normal) subgroup of B, and the map $B \to C$ is surjective with kernel A, so that we have $B/A \xrightarrow{\sim} C$.

We digress to define commutativity of diagrams. A diagram

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ {\scriptstyle\varphi}\downarrow & & \downarrow{\scriptstyle\psi} \\ C & \xrightarrow{g} & D \end{array}$$

is said to be *commutative* if $\psi \circ f = g \circ \varphi$. In general, a diagram consisting of abelian groups and homomorphisms between them is said to be commutative if any two 'paths' in the diagram that start at the same point in the diagram and end at the same point as well, define the same map between the groups standing at those points. (To make this into a rigorous definition, one first needs to define what a 'diagram' is; this may be done later.)

One can embed any homomorphism $B \xrightarrow{g} C$ in a unique way (up to isomorphism) in an exact sequence $0 \to A \to B \xrightarrow{g} C \to D \to 0$, namely by taking $A = \ker g$ and D equal to the *cokernel* $\operatorname{cok} g = C/g(B)$ of g, the maps $A \to B$ and $C \to D$ being the inclusion map and the canonical map, respectively. This exact sequence is called the *kernel-cokernel sequence.* The uniqueness statement means that for any exact sequence $0 \to A \to B \xrightarrow{g} C \to D \to 0$ there are unique isomorphisms $A \xrightarrow{\sim} \ker g$ and $D \xrightarrow{\sim} \operatorname{cok} g$ for which the diagram

$$\begin{array}{ccccccccccc}
0 & \longrightarrow & A & \longrightarrow & B & \longrightarrow & C & \longrightarrow & D & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \ker g & \longrightarrow & B & \longrightarrow & C & \longrightarrow & \operatorname{cok} g & \longrightarrow & 0
\end{array}$$

is commutative; here the vertical arrows starting at A and D are the isomorphisms, and those starting at B and C are the identity maps. Note that if B and C are finite we have the relation $\#A \cdot \#C = \#B \cdot \#D$ (see Exercise 4.2).

There are many 'diagram lemmas', which make assertions about exact sequences that are combined into commutative diagrams. As an example we present the most important one, which is called the *snake lemma.*

Snake lemma. *Any commutative diagram with exact rows*

$$\begin{array}{ccccccc}
 & & P & \longrightarrow & Q & \longrightarrow & R & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & X & \longrightarrow & Y & \longrightarrow & Z & &
\end{array}$$

can be embedded in a commutative diagram

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & A & \xrightarrow{f} & B & \xrightarrow{g} & C & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & P & \xrightarrow{m} & Q & \longrightarrow & R & \longrightarrow & 0 \\
 & & {\scriptstyle\alpha}\downarrow & & {\scriptstyle\beta}\downarrow & & \downarrow & & \\
0 & \longrightarrow & X & \longrightarrow & Y & \longrightarrow & Z & & \\
 & & {\scriptstyle\varphi}\downarrow & & \downarrow & & \downarrow & & \\
 & & D & \xrightarrow{i} & E & \xrightarrow{j} & F & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array}$$

with exact rows and columns; in addition, there is a homomorphism $C \to D$ such that the sequence $A \to B \to C \to D \to E \to F$ is exact. Also, if $0 \to P \to Q$ is exact, then $0 \to A \to B$ is exact; and if $Y \to Z \to 0$ is exact then $E \to F \to 0$ is exact; and if both conditions are satisfied, then $0 \to A \to B \to C \to D \to E \to F \to 0$ is exact.

Remark. Drawing the arrow $C \to D$ into the diagram one sees the 'snake' appear.

Proof. We first construct exact sequences $A \to B \to C$ and $D \to E \to F$. It is obvious that we can construct f, g, i, and j as induced homomorphisms. For example, treating A and B as the subgroups $\ker\alpha$ and $\ker\beta$ of P and Q, respectively, we define f as the restriction of m to A. By commutativity of $PQYX$, we have $mA \subset \ker\beta$, so f maps A to B and is a homomorphism.

What requires proof is $\operatorname{im} f = \ker g$ and $\operatorname{im} i = \ker j$. This is done by "diagram chasing". We illustrate this on the latter case. By exactness, X maps to 0 in Z, and hence to 0 in F. By commutativity, it follows that we have $j \circ i \circ \varphi = 0$ (the zero map). But $\varphi X = D$, so we have $(j \circ i)(D) = 0$. Thus $\operatorname{im} i \subset \ker j$. To show the opposite inclusion, let $e \in \ker j \subset E$. Because $Y \to E$ is onto, there is $y \in Y$ mapping to e. Let $y \mapsto z \in Z$. From $y \mapsto e \mapsto 0 \in F$ we see, by commutativity of $YZFE$, that z maps to $0 \in F$, i. e., z is in the kernel of the map $Z \to F$. Some $r \in R$ maps to z and some $q \in Q$ maps to r, by exactness of $R \to Z \to F$ and $Q \to R \to 0$, respectively. Let $q \mapsto y' \in Y$. By commutativity of $QRZY$ we must have $y' \mapsto z$. Now define $y'' = y - y'$. Firstly, $y'' \mapsto 0 \in Z$ so y'' belongs to the kernel of the map $Y \to Z$. Hence some $x \in X$ maps to y''. Secondly, $Q \to Y \to E$

is exact so $y' \mapsto 0 \in E$ and therefore $y'' \mapsto e$. By commutativity of $XYED$, we have $e = (i \circ \varphi)(x) \in (i \circ \varphi)X = i(\varphi X) = iD = \operatorname{im} i$, giving us the reverse inclusion.

Note that the proof of exactness at E that we just gave depends on the 0 in $Q \to R \to 0$. Likewise, the proof of exactness at B—which we leave to the reader—depends on the 0 in $0 \to X \to Y$.

To construct $C \to D$, we use an additional lemma. Given a commutative diagram

$$\Sigma : \quad \begin{tikzcd} A \arrow[r, "f"] \arrow[d, "h"'] \arrow[dr, "j"] & B \arrow[d, "g"] \\ C \arrow[r, "i"'] & D \end{tikzcd}$$

we define $\ker \Sigma = \ker j/(\ker f + \ker h)$ and $\operatorname{cok} \Sigma = (\operatorname{im} g \cap \operatorname{im} i)/\operatorname{im} j$.

Lemma. *Let*

$$\begin{tikzcd} A \arrow[r, "a"] \arrow[d, "c"'] \arrow[dr, phantom, "\Sigma"] & B \arrow[r, "b"] \arrow[d, "f"'] \arrow[dr, phantom, "\mathrm{T}"] & C \arrow[d, "d"'] \\ D \arrow[r, "e"] & E \arrow[r, "g"] & F \end{tikzcd}$$

be a commutative diagram with exact rows. Then f induces an isomorphism

$$\ker \mathrm{T} \xrightarrow{\sim} \operatorname{cok} \Sigma.$$

Proof. Observe that $\ker gf \xrightarrow{f} \ker g$ is a homomorphism. By the appropriate isomorphism theorem, we have $\ker gf/\ker f \xrightarrow{\sim} \operatorname{im} f \cap \ker g = \operatorname{im} f \cap \operatorname{im} e$ (this last equality is true by exactness). From $\ker b = \operatorname{im} a$ it follows that this mapping sends the image $(\ker f + \ker b)/\ker f$ of $\ker b$ in $\ker gf/\ker f$ onto the subgroup $\operatorname{im} fa$ of $\operatorname{im} f \cap \operatorname{im} e$. Factoring out corresponding subgroups—as we can do in abelian groups—we obtain

$$\ker gf/(\ker f + \ker b) \xrightarrow{\sim} (\operatorname{im} f \cap \operatorname{im} e)/\operatorname{im} fa.$$

In other words, we have $\ker \mathrm{T} \xrightarrow{\sim} \operatorname{cok} \Sigma$. This proves the lemma.

To conclude the proof of the Snake lemma, we enlarge the diagram a bit:

$$\begin{tikzcd}
 & & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] \arrow[r] \arrow[dr, phantom, "\Sigma_1"] & \operatorname{cok} g \arrow[d] & \\
 & & A \arrow[d] \arrow[r] & B \arrow[d] \arrow[r, "g"] \arrow[dr, phantom, "\Sigma_2"] & C \arrow[d] \arrow[r] \arrow[dr, phantom, "\mathrm{T}_1"] & \operatorname{cok} g \arrow[d] \arrow[r] & 0 \\
 & & P \arrow[d] \arrow[r] \arrow[dr, phantom, "\Sigma_3"] & Q \arrow[d] \arrow[r] \arrow[dr, phantom, "\mathrm{T}_2"] & R \arrow[d] \arrow[r] & 0 & \\
 & 0 \arrow[d] \arrow[r] \arrow[dr, phantom, "\Sigma_4"] & X \arrow[d] \arrow[r] \arrow[dr, phantom, "\mathrm{T}_3"] & Y \arrow[d] \arrow[r] & Z \arrow[d] & & \\
0 \arrow[r] & \ker i \arrow[d] \arrow[r] \arrow[dr, phantom, "\mathrm{T}_4"] & D \arrow[d] \arrow[r, "i"] & E \arrow[d] \arrow[r] & F \arrow[d] & & \\
 & \ker i \arrow[r] & 0 & 0 & 0. & &
\end{tikzcd}$$

Apply the lemma to the pairs (Σ_1, T_1), (Σ_2, T_1), (Σ_2, T_2), $\ldots$, (Σ_4, T_4), to deduce

$$\operatorname{cok} g \cong \operatorname{cok} \Sigma_1 \cong \ker \mathrm{T}_1 \cong \operatorname{cok} \Sigma_2 \cong \ldots \cong \operatorname{cok} \Sigma_4 \cong \ker \mathrm{T}_4 \cong \ker i.$$

Hence $A \to B \to C \to D \to E \to F$ is exact.

If $0 \to P \to Q$ is exact then m is injective, and since we obtained the map $A \to B$ by restricting m to A (when viewed as a subgroup of P) it is injective as well. A similar argument shows that exactness of $Y \to Z \to 0$ implies exactness of $E \to F \to 0$. The last assertion of the Snake lemma is now obvious. This completes the proof.

A mistaken sense of symmetry might lead one to think that $B \to C \to 0$ and $0 \to D \to E$ are also exact, especially in the case $0 \to P \to Q$ and $Y \to Z \to 0$ are exact. This need not be true; but the groups $\operatorname{cok} g$ and $\ker i$, which measure the failure for it to be true, are isomorphic, as we saw in the proof. Also, this 'obstruction group' $\operatorname{cok} g \cong \ker i$ can be read from the initially given diagram, since as we saw it is the same as $\ker \mathrm{T}_2 \cong \operatorname{cok} \Sigma_3$.

4. Exercises

4.1. Can you generalize the results of this section to not necessarily abelian groups, with supplemental conditions relating to the normality of certain subgroups?

4.2. (a) Let

$$0 \to A_1 \to A_2 \to A_3 \to \ldots \to A_{n-1} \to A_n \to 0$$

be an exact sequence of finite abelian groups. Prove that

$$\prod_{\substack{1 \le i \le n \\ i \text{ even}}} \#A_i = \prod_{\substack{1 \le i \le n \\ i \text{ odd}}} \#A_i.$$

(b) Formulate and prove a similar result when finite abelian groups are replaced by finite-dimensional vector spaces over a field k, and all maps are supposed to be k-linear.

4.3. The *cokernel* $\operatorname{cok} f$ of a homomorphism $f\colon A \to B$ of abelian groups is defined to be B/fA. Let $f\colon A \to B$ and $g\colon B \to C$ be two homomorphisms of abelian groups. Prove that there is an exact sequence

$$0 \to \ker f \to \ker gf \to \ker g \to \operatorname{cok} f \to \operatorname{cok} gf \to \operatorname{cok} g \to 0.$$

Can you deduce this from the snake lemma?

4.4. Let n be an integer, and let A be an additively written abelian group. Write n_A for the endomorphism of A that maps every $a \in A$ to na. If $\ker n_A$ and $\operatorname{cok} n_A$ are finite, then we define $h_n(A) = (\#\ker n_A)/\#\operatorname{cok} n_A$, and we say that $h_n(A)$ is *defined.*

(a) Let $0 \to A \to B \to C \to 0$ be an short exact sequence of abelian groups, and suppose that two of $h_n(A)$, $h_n(B)$, $h_n(C)$ are defined. Prove that the third one is also defined, and that one has $h_n(B) = h_n(A) \cdot h_n(C)$.

(b) Compute $h_n(A)$ for $A = \mathbf{Z}$, $\mathbf{Q}$, $\mathbf{R}/\mathbf{Z}$, and when A is any finite abelian group. (Do not use the structure theorem for finite abelian groups.)

4.5. Let the notation be as in Exercise 4.4.

(a) Let n and m be integers, and let A be an abelian group. Prove that $h_{nm}(A)$ is defined if and only if $h_n(A)$ and $h_m(A)$ are both defined, and that one has $h_{nm}(A) = h_n(A) \cdot h_m(A)$ if all three are defined.

(b) Let n be an integer. Determine the set of all numbers that occur as $h_n(A)$ for some abelian group A for which $h_n(A)$ is defined.

Modules. Let R be a ring. An *R-module* or *left R-module* is an abelian group M together with a map $R \times M \to M$, $(r, m) \mapsto rm$, satisfying the following four rules for all r, $r' \in R$ and m, $m' \in M$:

$$r(m + m') = rm + rm', \quad (r + r')m = rm + r'm, \quad (rr')m = r(r'm), \quad 1m = m,$$

where 1 denotes the unit element of R. If the third rule is replaced by $(rr')m = r'(rm)$ one obtains the definition of a *right* R-module; in this case one usually writes mr instead of rm, so that the third rule assumes the more natural appearance $m(rr') = (mr)r'$. If M, M' are R-modules, then an *R-homomorphism* or *R-linear map* $M \to M'$ is a group homomorphism $f: M \to M'$ satisfying $f(rm) = r.f(m)$ for all $r \in R$, $m \in M$.

4.6. A *differential group* is an abelian group A equipped with an endomorphism $d = d_A$ satisfiying $d^2 = 0$. If A and B are differential groups, then a *morphism* from A to B is a group homomorphism $f: A \to B$ satisfying $f d_A = d_B f$.

(a) Prove that the notion of a differential group is equivalent to the notion of a $\mathbf{Z}[X]/(X^2)$-module, in such a manner that the morphisms correspond to the $\mathbf{Z}[X]/(X^2)$-linear maps. (*Note*: a more rigorous formulation of this exercise can be given in the language of categories, which shall be introduced later.)

(b) Construct, for every differential group A, a group $H(A)$ that fits into an exact sequence $0 \to H(A) \to \operatorname{cok} d_A \to \ker d_A \to H(A) \to 0$, the middle arrow being induced by d_A. The group $H(A)$ is called the *homology group* of A.

4.7. Let $0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$ be an exact sequence of differential groups, the maps being morphisms of differential groups (see Exercise 4.6). Construct an exact triangle

$$\begin{array}{ccc} H(A) & \to & H(B) \\ & \nwarrow & \downarrow \\ & & H(C). \end{array}$$

4.8. A *complex* is a sequence $(A_n)_{n \in \mathbf{Z}}$ of abelian groups together with a sequence $(d_n)_{n \in \mathbf{Z}}$ of homomorphisms $d_n: A_n \to A_{n+1}$ such that for all $n \in \mathbf{Z}$ one has $d_n d_{n-1} = 0$. The nth *homology group* $H_n(A)$ of such a complex A is defined by $H_n(A) = (\ker d_n)/d_{n-1}A_{n-1}$.

Suppose $A = \big((A_n)_{n \in \mathbf{Z}}, (d_n)_{n \in \mathbf{Z}}\big)$, $B = \big((B_n)_{n \in \mathbf{Z}}, (d_n)_{n \in \mathbf{Z}}\big)$, $C = \big((C_n)_{n \in \mathbf{Z}}, (d_n)_{n \in \mathbf{Z}}\big)$ are three complexes, and that for every n one has a short exact sequence $0 \to A_n \xrightarrow{f_n} B_n \xrightarrow{g_n} C_n \to 0$ such that for all n one has $f_n d_{n-1} = d_{n-1} f_{n-1}$ and $g_n d_{n-1} = d_{n-1} g_{n-1}$. Construct group homomorphisms $H_{n-1}(C) \to H_n(A)$ that together with the natural maps $H_n(A) \to H_n(B)$ and $H_n(B) \to H_n(C)$ induced by f_n and g_n yield a long exact sequence

$$\cdots \to H_{n-1}(C) \to H_n(A) \to H_n(B) \to H_n(C) \to H_{n+1}(A) \to \cdots.$$

The homomorphisms $H_{n-1}(C) \to H_n(A)$ are called the *connecting* homomorphisms.

4.9. Let R be a ring. Verify that the results of this section remain valid if one works with R-modules instead of abelian groups, and with R-linear maps instead of homomorphisms.

5. Free groups and presentations

Let S be a set. We shall construct a group $F(S)$ containing S with the following "universal" property: *for any group G and any (set-theoretic) map $f\colon S \to G$, there is a unique group homomorphism $g\colon F(S) \to G$ such that $g|S = f$.* In other words, if $j\colon S \to F(S)$ denotes the inclusion map, then for any map $f\colon S \to G$ there is a unique group homomorphism g making the diagram

$$\begin{tikzcd} S \arrow[r, "j"] \arrow[dr, "f"'] & F(S) \arrow[d, "g"] \\ & G \end{tikzcd}$$

commutative. The universal property may be succinctly expressed by saying that for any group G the map

$$j^*\colon \mathrm{Hom}(F(S), G) \to G^S, \qquad g \mapsto f = g|S = gj$$

from the set $\mathrm{Hom}(F(S), G)$ of group homomorphisms $F(S) \to G$ to the set G^S of all maps $S \to G$ is *bijective.* As we shall see, this property characterizes $F(S)$ up to a canonical isomorphism; it will be called the *free group* on S. The word "free" is intended to convey the idea that one builds $F(S)$ up from the elements of S without introducing any "relations" between them; indeed, any relation that the elements of S would satisfy in $F(S)$ they would satisfy in *any* group G, by the universal property.

One readily verifies that for $S = \emptyset$ the trivial group $F(S) = \{1\}$ has the universal property, and that for $S = \{a\}$ one can take $F(S) = \{a^n : n \in \mathbf{Z}\}$, with multiplication $a^n \cdot a^m = a^{n+m}$; in the latter example, the "symbols" a^n are viewed as pairwise different, so that $F(S)$ is just an isomorphic copy of $\mathbf{Z}$.

Construction of F(S). Define the *alphabet* to be the disjoint union $S \cup \bar{S}$, where $\bar{S} = \{\bar{s} : s \in S\}$ is a copy of S (formally, one may define the alphabet to be $S \times \{1, -1\}$, identifying s with $(s, 1)$ and $\bar{s}$ with $(s, -1)$, for $s \in S$; the element $\bar{s}$ is going to be the inverse of s in the group $F(S)$). For $s \in S$, we shall write $\bar{\bar{s}} = s$, so that $^-$ is defined on all of $S \cup \bar{S}$. A finite sequence $s_0 s_1 \cdots s_{t-1}$ of elements of the alphabet is called a *word.* (Formally, a word is a pair (t, f), where $t \in \mathbf{Z}$, $t \geq 0$, and $f\colon \{0, 1, \ldots, t-1\} \to S \cup \bar{S}$ is a map.) The empty sequence, with $t = 0$, is also a word. A *reduced word* is a word for which there exists no i with $s_{i+1} = \bar{s}_i$. For example, if $S = \{a\}$, then any reduced word is of the form $aa \cdots a$ or $\bar{a}\bar{a} \cdots \bar{a}$, or it is the empty sequence. We now define $F(S)$ to be the set of all reduced words, with multiplication $*$ given by:

$$s_0 s_1 \cdots s_{t-1} * r_0 r_1 \cdots r_{u-1} = s_0 \cdots s_{t-j-1} r_j \cdots r_{u-1}$$

where j is the largest number less than or equal to $\min\{u,t\}$ such that $r_i = \bar{s}_{t-i-1}$ for all $i \in \{0, 1, \ldots, j-1\}$; this guarantees that the result is again reduced. In the event that $j = u$ or $j = t$, all r_i or all s_i disappear.

Theorem 1. *For any set S, the set $F(S)$ with the multiplication $*$ just defined is a group, and it has the universal property formulated above.*

Proof. One can directly check the group axioms, but the proof of associativity is a bit painful. It can be avoided by the following trick, which is due to Van der Waerden. For $s \in S \cup \bar{S}$ define the map $\sigma_s\colon F(S) \to F(S)$ by $\sigma_s(w) = s * w$. That is, if $w = r_0 r_1 \cdots r_{u-1}$, then

$$\sigma_s(w) = \begin{cases} r_1 \cdots r_{u-1}, & \text{if } u \geq 1 \text{ and } r_0 = \bar{s}; \\ s r_0 \cdots r_{u-1}, & \text{otherwise.} \end{cases}$$

One easily checks that $\sigma_{\bar{s}}\sigma_s = 1_{F(S)}$, and applying this result to $\bar{s}$ in the role of s one finds also $\sigma_s\sigma_{\bar{s}} = 1_{F(S)}$. Hence $\sigma_{\bar{s}}$ is a two-sided inverse of σ_s, so one has $\sigma_s \in \operatorname{Sym} F(S)$. Next, define the map $\varphi\colon F(S) \to \operatorname{Sym} F(S)$ by $\varphi(s_0 \cdots s_{t-1}) = \sigma_{s_0} \cdots \sigma_{s_{t-1}}$. From $\sigma_s\sigma_{\bar{s}} = 1_{F(S)}$ and the definition of $*$ one deduces that $\varphi(w_1 * w_2) = \varphi(w_1)\varphi(w_2)$. Also, one has $\varphi(s)^{-1} = \varphi(\bar{s})$. One deduces that $\varphi F(S)$ is a subgroup of $\operatorname{Sym} F(S)$. The map $\psi\colon \operatorname{Sym} F(S) \to F(S)$ sending σ to $\sigma(1)$, where $1 \in F(S)$ denotes the empty sequence, satisfies $\psi\varphi = 1_{F(S)}$, so φ is a bijection from $F(S)$ to $\varphi F(S)$. The latter is a group, so by "transport of structure"—i. e., by applying ψ—one deduces that $F(S)$ is a group as well.

The proof of the universal property is immediate. Given f, one defines $g\colon F(S) \to G$ by $g(s_0 \cdots s_{t-1}) = f(s_0) \cdots f(s_{t-1})$, where we write $f(\bar{s}) = f(s)^{-1}$ for $s \in S$. Clearly g is a group homomorphism $F(S) \to G$ that extends f, and it is the only such group homomorphism. This proves the theorem.

The pair consisting of the group $F(S)$ and the inclusion map $j\colon S \to F(S)$ is, up to a unique isomorphism, uniquely determined by the universal property. This is the content of the following theorem.

Theorem 2. *Let S be a set, E a group, and $i\colon S \to E$ a map. Suppose that E and i satisfy the universal property; that is, for every group G the map $i^*\colon \operatorname{Hom}(E, G) \to G^S$ sending g to gi is bijective. Then there is a unique group isomorphism $\alpha\colon F(S) \to E$ satisfying $i = \alpha j$.*

Proof. By the universal property of $F(S)$, applied to $G = E$ and $f = i$, there is a unique group homomorphism $\alpha\colon F(S) \to E$ such that $i = \alpha j$. To prove the theorem, it suffices to show that α is an isomorphism. We shall do so by exhibiting a two-sided inverse. By the universal property of E, applied to $G = F(S)$, there is a unique group homomorphism $\beta\colon E \to F(S)$ such that $\beta i = j$. The maps 1_E, $\alpha\beta \in \operatorname{Hom}(E, E)$ have the same image under the map $i^*\colon \operatorname{Hom}(E, E) \to G^S$, since $i^*(1_E) = 1_E i = i$ and $i^*(\alpha\beta) = \alpha\beta i = \alpha j = i$;

since i^* is bijective it follows that $1_E = \alpha\beta$. The same argument, with $F(S)$ in the role of E, shows that $1_{F(S)} = \beta\alpha$. Hence β is indeed a two-sided inverse of α, as required. This proves Theorem 2. The proof incidentally shows that $i = \alpha j$ is automatically injective, since both α and j are.

The argument just given allows an important generalization. Let two groups E and F be fixed, and suppose that for each group G one has a bijection $\eta_G\colon \mathrm{Hom}(E, G) \to \mathrm{Hom}(F, G)$; above, this was actually the case (with $F = F(S)$), since $\eta_G = j^{*-1}i^*$ is a bijection $\mathrm{Hom}(E, G) \xrightarrow{\sim} G^S \xrightarrow{\sim} \mathrm{Hom}(F, G)$. Such a system (η_G) of bijections, with G ranging over all groups, is called *functorial in* G if for all groups G and H and all homomorphisms $h\colon G \to H$ the diagram

$$\begin{array}{ccc} \mathrm{Hom}(E,G) & \xrightarrow{\eta_G} & \mathrm{Hom}(F,G) \\ {\scriptstyle h_*}\big\downarrow & & {\scriptstyle h_*}\big\downarrow \\ \mathrm{Hom}(E,H) & \xrightarrow{\eta_H} & \mathrm{Hom}(F,H) \end{array}$$

is commutative; here h_* is the map $g \mapsto h \circ g$. In other words, functoriality means that for all homomorphisms $h\colon G \to H$ we have $h_* \circ \eta_G = \eta_H \circ h_*$. By abuse of language, it is often simply said that the bijection η_G is functorial in G, even though the entire system of bijections is meant. We note that the maps $\eta_G = j^{*-1}i^*$ from Theorem 2 are indeed functorial in G. To prove this, remark that from $h_*(g) = h \circ g$ and $i^*(g) = g \circ i$ it evidently follows that $h_*i^* = i^*h_*$, and in the same way one proves that $h_*j^* = j^*h_*$; composing the latter equality with j^{*-1} on both sides one obtains $j^{*-1}h_* = h_*j^{*-1}$, and this leads to $h_* \circ \eta_G = h_*j^{*-1}i^* = j^{*-1}h_*i^* = j^{*-1}i^*h_* = \eta_H \circ h_*$, as required.

We can now generalize Theorem 2 by saying, in imprecise terms, that a group is determined, up to isomorphism, by the families of outgoing arrows to all other groups; the precise formulation in Theorem 3 involves the functoriality condition. Exercise 5.7(b) shows that this condition cannot be omitted.

Theorem 3. *Let E and F be groups. Suppose that for every group G one has a bijection $\eta_G\colon \mathrm{Hom}(E, G) \to \mathrm{Hom}(F, G)$, and that this system of bijections is functorial in G. Then one has $E \cong F$; more precisely, there is a unique group isomorphism $\varphi\colon F \to E$ such that for all groups G and all $f \in \mathrm{Hom}(E, G)$ one has $\eta_G(f) = f\varphi$.*

Proof. First observe that if φ is any map satisfying the conclusion of the theorem, then putting $G = E$ and $f = 1_E$ we see that we must have $\varphi = 1_E\varphi = \eta_E(1_E)$. Hence there can be at most one such φ; to show that there is at least one it suffices to define $\varphi = \eta_E(1_E)$ and to show that one has $\eta_G(f) = f\varphi$ for all G and f, and that φ is an isomorphism.

To prove the first statement, we use that by functoriality the diagram

$$\begin{array}{ccc}\mathrm{Hom}(E,E) & \xrightarrow{\eta_E} & \mathrm{Hom}(F,E)\\ {\scriptstyle f_*}\big\downarrow & & {\scriptstyle f_*}\big\downarrow\\ \mathrm{Hom}(E,G) & \xrightarrow{\eta_G} & \mathrm{Hom}(F,G)\end{array}$$

is commutative, so that indeed we have $\eta_G(f) = \eta_G\big(f_*(1_E)\big) = f_*\big(\eta_E(1_E)\big) = f\varphi$.

Since $\eta_F\colon \mathrm{Hom}(E,F) \to \mathrm{Hom}(F,F)$ is bijective, there is a unique $\psi \in \mathrm{Hom}(E,F)$ with $\eta_F(\psi) = 1_F$ or, equivalently, $\psi\varphi = 1_F$. To prove that φ is an isomorphism it suffices to check that $\varphi\psi = 1_E$. We have $\eta_E(\varphi\psi) = \varphi\psi\varphi = \varphi 1_F = \varphi = 1_E\varphi = \eta_E(1_E)$. Since η_E is bijective it follows that $\varphi\psi = 1_E$, as required. This concludes the proof of Theorem 3.

Let S be a set and let $F(S)$ be the free group on S. Take $r \in F(S)$. A map from S to a group G is said to *satisfy* r if r is in the kernel of the corresponding group homomorphism $F(S) \to G$; and it is said to satisfy a subset $R \subset F(S)$ if it satisfies all elements of R.

Let S be a set and $R \subset F(S)$. We define

$$\langle S\colon R\rangle = F(S)/N_R,$$

where N_R is the subgroup of $F(S)$ generated by $\{grg^{-1} : r \in R,\, g \in F(S)\}$; this is clearly a normal subgroup of $F(S)$, and it is in fact the smallest normal subgroup of $F(S)$ that contains R. One calls $\langle S\colon R\rangle$ the group *determined by the generators S and the relations R.*

Proposition 4. *For any group G, we have a functorial bijection*

$$\mathrm{Hom}(\langle S\colon R\rangle, G) \xrightarrow{\sim} \{\textit{maps } S \to G \textit{ satisfying } R\}.$$

Furthermore, this system of functorial bijections characterizes $\langle S\colon R\rangle$ up to a unique isomorphism.

Proof. Let c be the canonical homomorphism from $F(S)$ to $\langle S\colon R\rangle$, sending x to xN_R. By the first homomorphism theorem and the definition $\langle S\colon R\rangle = F(S)/N_R$, the map $c^*\colon \mathrm{Hom}(\langle S\colon R\rangle, G) \to \mathrm{Hom}(F(S), G)$ sending h to hc gives, for every group G, a bijection

$$\mathrm{Hom}(\langle S\colon R\rangle, G) \to \{f \in \mathrm{Hom}(F(S), G) : N_R \subset \ker f\},$$

and this bijection is functorial in G. Moreover, the set on the right is by definition of N_R equal to

$$\{f \in \mathrm{Hom}(F(S), G) : R \subset \ker f\}.$$

By the universal property of $F(S)$, this set in turn admits a bijection to

$$\{\text{maps } S \to G \text{ satisfying } R\},$$

and this bijection is again functorial in G. Composing these functorial bijections we obtain the functorial bijection that the proposition asserts to exist.

Suppose E is another group for which there is a functorial bijection as in the proposition. Composing this functorial bijection with the inverse of the other one we obtain a functorial bijection from $\mathrm{Hom}(E, G)$ to $\mathrm{Hom}(\langle S\colon R\rangle, G)$. By Theorem 3, it induces a unique isomorphism $E \to \langle S\colon R\rangle$. This proves Proposition 4.

If A is a group then a *presentation* of A is an isomorphism of A with a group of the form $\langle S\colon R\rangle$.

Proposition 5. *Any group A has a presentation.*

Proof. Choose a subset $S \subset A$ that generates A. (This can always be done, e. g. with $S = A$.) Then the inclusion map $S \to A$ induces a homomorphism $g\colon F(S) \to A$, and g is surjective. Choose a subset $R \subset \ker g$ with $N_R = \ker g$. (Again we can do this, e. g. with $R = \ker g$.) Then we have $\langle S\colon R\rangle \cong F(S)/\ker g \cong \operatorname{im} g = A$. This proves Proposition 5.

Alternative proof. Let $S = \{x_a : a \in A\}$, so that the set S is just a copy of the underlying set of A, and put $R = \{x_a x_b \bar{x}_{ab} : a, b \in A\} \subset F(S)$. Then for any group G we have

$$\begin{aligned}
\mathrm{Hom}(\langle S\colon R\rangle, G) &\cong \{\text{maps } S \to G \text{ satisfying } R\} \\
&= \{\text{maps } \varphi\colon S \to G : \forall a, b \in A : \varphi(x_a)\varphi(x_b) = \varphi(x_{ab})\} \\
&\cong \{\text{maps } \psi\colon A \to G : \forall a, b \in A : \psi(a)\psi(b) = \psi(ab)\} \\
&= \mathrm{Hom}(A, G).
\end{aligned}$$

Each of the bijections is functorial in G, so by Theorem 3 we have $A \cong \langle S\colon R\rangle$, as required.

A presentation $\langle S\colon R\rangle$ of a group is called *finite* if both S and R are finite. If $S = \{s_1, \ldots, s_n\}$ and $R = \{r_1, \ldots, r_m\}$, we also write $\langle S\colon R\rangle = \langle s_1, \ldots, s_n : r_1, \ldots, r_m\rangle$. Often, one writes $r_i = 1$ instead of just r_i; for example, instead of $\langle x, y : x^2y^{-2},\ xyxy^{-1}\rangle$ one may write $\langle x, y : x^2y^{-2} = 1,\ xyxy^{-1} = 1\rangle$, or even $\langle x, y : x^2 = y^2,\ xyx = y\rangle$. A group that has a finite presentation is called *finitely presented.*

Examples of finite presentations: $\langle x : x^7\rangle = \langle x : x^7 = 1\rangle$ is a cyclic group of order 7, and $\langle x : \rangle$ is an infinite cyclic group.

As the proof of Proposition 5 shows, any group has many presentations. One often prefers one that either reflects the way in which the group is built up, or one with a small number of generators and relations. For example, consider the group S_3. It has the presentation $\langle x, y : y^3 = 1,\ x^2 = 1,\ xyx^{-1} = y^{-1}\rangle$, which reflects that $S_3 \cong C_3 \rtimes C_2$

(with respect to a certain action of C_2 on C_3). A more economical presentation is $\langle x, y : x^2 = 1,\ xyx^{-1} = y^2\rangle$; note that, in the last group, one has $y = x^2yx^{-2} = x(xyx^{-1})x^{-1} = xy^2x^{-1} = (xyx^{-1})^2 = y^4$, and therefore $y^3 = 1$.

Sums of groups. Let A and B be groups, and choose presentations $A \cong \langle S\colon R\rangle$ and $B \cong \langle T : Q\rangle$ with $S \cap T = \emptyset$. The *sum* (or *free product* in the older literature) $A \circ B$ of A and B is defined by $A \circ B = \langle S \cup T : R \cup Q\rangle$. To prove that this is independent of the presentations chosen, we note that for any group G we have

$$\begin{aligned} \mathrm{Hom}(A \circ B, G) &\cong \{\text{maps } S \cup T \to G \text{ satisfying } R \cup Q\} \\ &\cong \{\text{maps } S \to G \text{ satisfying } R\} \times \{\text{maps } T \to G \text{ satisfying } Q\} \\ &\cong \mathrm{Hom}(A, G) \times \mathrm{Hom}(B, G), \end{aligned}$$

and these bijections are functorial in G. Thus, if different presentations would give rise to, say, a group $A \bullet B$, then one would obtain functorial bijections

$$\mathrm{Hom}(A \circ B, G) \cong \mathrm{Hom}(A, G) \times \mathrm{Hom}(B, G) \cong \mathrm{Hom}(A \bullet B, G),$$

so Theorem 3 would yield an isomorphism $A \circ B \cong A \bullet B$.

Let the bijection $\mathrm{Hom}(A \circ B, A \circ B) \xrightarrow{\sim} \mathrm{Hom}(A, A \circ B) \times \mathrm{Hom}(B, A \circ B)$, which one obtains by putting $G = A \circ B$, send the identity map $1_{A\circ B}$ to the pair (i, j), where $i\colon A \to A \circ B$ and $j\colon B \to A \circ B$ are group homomorphisms. Then the bijection $\mathrm{Hom}(A \circ B, G) \xrightarrow{\sim} \mathrm{Hom}(A, G) \times \mathrm{Hom}(B, G)$ maps f to (fi, fj), for any group G. The bijectivity of this map is equivalent to the following *universal property* of $A \circ B$: if G is any group, and $g\colon A \to G$, $h\colon B \to G$ are arbitrary group homomorphisms, then there is a unique group homomorphism $f\colon A \circ B \to G$ for which $g = fi$ and $h = fj$.

In Exercise 5.1 one finds an explicit description of the elements of $A \circ B$. It shows in particular that $A \circ B$ contains (isomorphic copies of) A and B as *subgroups*. If $(A_i)_{i\in I}$ is any system of groups, with a possibly infinite index set I, one can in a similar manner define their sum $\bigcirc_{i\in I} A_i$ in such a way that one has bijections

$$\mathrm{Hom}\big(\bigcirc_{i\in I} A_i, G\big) \cong \prod_{i\in I} \mathrm{Hom}(A_i, G)$$

that are functorial in G. Exercise 5.1 carries over to this situation as well.

Let A and B be groups. The pair of group homomorphisms $A \to A \times B$, $B \to A \times B$ sending $a \in A$ to $(a, 1)$ and $b \in B$ to $(1, b)$, respectively, combines into a surjective group homomorphism $A \circ B \to A \times B$. However, this group homomorphism is only rarely injective (see Exercise 5.2), so the sum is really different from the product. In intuitive terms, this is because in the product group $A \times B$ the elements coming from A commute with those coming from B, which is expressed by relations that do not occur in the definition of the sum group.

On the other hand, it is true that the product $A \times B$ has a universal property similar to the sum, but not with respect to *outgoing* arrows but with respect to *incoming* arrows. That is, if $\pi_1 \colon A \times B \to A$ and $\pi_2 \colon A \times B \to B$ are the coordinate projections, then for any group G one has a bijection

$$(\pi_{1*}, \pi_{2*}) \colon \mathrm{Hom}(G, A \times B) \to \mathrm{Hom}(G, A) \times \mathrm{Hom}(G, B), \quad f \mapsto (\pi_1 f, \pi_2 f);$$

also, this bijection is functorial in G, which now means that for any group homomorphism $h \colon G \to H$ the two maps $(\pi_{1*}, \pi_{2*}) \circ h^*$ and $(h^*, h^*) \circ (\pi_{1*}, \pi_{2*})$ from $\mathrm{Hom}(H, A \times B)$ to $\mathrm{Hom}(G, A) \times \mathrm{Hom}(G, B)$ are the same (with $h^*(f) = fh$). This universal property again characterizes $A \times B$, by a theorem similar (but "dual") to Theorem 3 (see Exercise 5.6). The situation is best understood from the point of view of categories (see Section 7, and Exercises 7.22 and 7.24).

If we restrict to *abelian* groups—not just for A and B, but also for the "test groups" G—then the product $A \times B$ has *both* universal properties. That is, if A and B are abelian groups, then for all abelian groups G one has functorial group isomorphisms

$$(\pi_{1*}, \pi_{2*}) \colon \mathrm{Hom}(G, A \times B) \to \mathrm{Hom}(G, A) \times \mathrm{Hom}(G, B), \quad f \mapsto (\pi_1 f, \pi_2 f),$$
$$\pi_1^* + \pi_2^* \colon \mathrm{Hom}(A, G) \times \mathrm{Hom}(B, G) \to \mathrm{Hom}(A \times B, G), \quad (f, g) \mapsto f\pi_1 + g\pi_2;$$

remember that $\mathrm{Hom}(A, G)$ is a group for abelian G. When A and B are abelian, one often writes $A \oplus B$ instead of $A \times B$, and refers to it as the *direct sum* of A and B.

5. Exercises

5.1. Let A, B be two groups.

(a) Imitating the construction of the free group, construct a group $A \bullet B$ that contains isomorphic copies of A and B as subgroups intersecting only in 1, with the property that every element of $A \bullet B$ has a unique representation as an alternating word $x_0 x_1 \cdots x_{t-1}$ with $t \geq 0$, $x_i \in A \cup B$, $x_i \neq 1$; *alternating* means that there is no i with $\{x_i, x_{i+1}\} \subset A$ or $\{x_i, x_{i+1}\} \subset B$.

(b) Exhibit for each group G a bijection $\mathrm{Hom}(A \bullet B, G) \to \mathrm{Hom}(A, G) \times \mathrm{Hom}(B, G)$ that is functorial in G, and prove that $A \bullet B \cong A \circ B$.

5.2. Let A, B be two groups. Show that the map $A \circ B \to A \times B$ defined in the text is an isomorphism if and only if at least one of the groups A, B is trivial.

5.3. Let S be a set with $\#S = 2$. Use the universal properties of $\mathbf{Z} \circ \mathbf{Z}$ and the free group $F(S)$ on S to show that $\mathbf{Z} \circ \mathbf{Z} \cong F(S)$.

5.4. Let C_2 be a group of order 2, and let $G = C_2 \circ C_2$.

(a) Prove that $H = \{x \in G : x \text{ does not have order } 2\}$ is a subgroup of index 2 of G, that H is infinite cyclic, and that $G \cong H \rtimes \operatorname{Aut} H$.

(b) Prove that G is isomorphic to the subgroup of $\operatorname{Sym} \mathbf{Z}$ generated by the two permutations $x \mapsto -x$ and $x \mapsto 1 - x$ of $\mathbf{Z}$. To which subgroup of $\operatorname{Sym} \mathbf{Z}$ does the subgroup H from (a) correspond?

5.5. Let G be as in Exercise 5.4.

(a) Determine all subgroups of G. Which of them are normal, and to which known groups are the corresponding factor groups isomorphic?

(b) Classify all groups that can be generated by two distinct elements of order 2. How many such groups are there (up to isomorphism) that are infinite?

5.6. Let E, F be groups with the property that for all groups G one has a functorial bijection $\eta_G\colon \operatorname{Hom}(G, E) \to \operatorname{Hom}(G, F)$. Prove that there is an isomorphism $\varphi\colon E \to F$ such that for all groups G and all $f \in \operatorname{Hom}(G, E)$ one has $\eta_G(f) = \varphi \circ f$.

5.7. For this exercise you need some knowledge of set theory.

(a) Prove that for every group G there is a bijection $\operatorname{Hom}(G, \mathbf{Z}) \to \operatorname{Hom}(G, \mathbf{Z} \times \mathbf{Z})$, but that there does not exist a functorial system of such bijections.

(b) Prove that for every group G there is a bijection $\operatorname{Hom}(\mathbf{Q}, G) \to \operatorname{Hom}(\mathbf{Q} \circ \mathbf{Q}, G)$, but that there does not exist a functorial system of such bijections.

5.8. Let G be a group and N a normal subgroup of G such that $G/N \cong F(S)$ for some set S. Prove: $G \cong N \rtimes F(S)$ with respect to some group homomorphism $F(S) \to \operatorname{Aut} N$.

5.9. (a) Prove that any group of the form $\langle x_1, x_2, \ldots, x_n : r_1, \ldots, r_m \rangle$ with $m < n$ is infinite.

(b) Prove that the group $\langle x, y : xx = yy,\ xyx = y \rangle$ is finite. To which group that you know is it isomorphic?

5.10. Let the group G be defined with generators x_n (for $n \in \mathbf{Z}$, $n \geq 1$) and relations $x_n^n = x_{n-1}$ (for $n \in \mathbf{Z}$, $n > 1$). To which group that you know is G isomorphic?

5.11. Find a presentation for the Klein four group V_4, for the quaternion group Q of order 8, for the groups $\mathbf{Z} \times \mathbf{Z}$ and $\mathbf{Q}/\mathbf{Z}$, and for the additive group $\mathbf{R}$ of real numbers. Which of these groups have *finite* presentations?

5.12. Show that each finite group is finitely presented.

5.13. Let A be a finite abelian group, and let $a_1, a_2, \ldots, a_t \in A$ be a finite system of generators for A. Prove that A has a presentation $A = \langle x_1, x_2, \ldots, x_t : r_1, \ldots, r_m \rangle$ where the x_i correspond to the a_i and the number m of relations equals $t(t+1)/2$.

5.14. Does the Klein four group V_4 have a presentation with *two* relations? Exhibit one, or show that it does not exist. (This is not so easy. It helps to construct a group G of order 32 that can be generated by two elements, with the properties that $G/Z(G) \cong V_4$ and $z^2 = 1$ for each $z \in Z(G)$.)

5.15. For a positive integer n, let G_n be the group with generators $s_0, \ldots, s_{n-1}$ and relations $s_{i+1}s_is_{i+1}^{-1} = s_i^2$ (for $0 \le i < n$), with $s_n = s_0$.

(a) Prove that the groups G_1, G_2, G_3 are trivial. *Note*: one can show that G_4 is non-trivial.

(b) Let n be a positive integer and let H be a finite group. Prove that every group homomorphism $G_n \to H$ is trivial. (*Hint*: first prove that $n = 1$ is the only positive integer n for which n divides $2^n - 1$.)

5.16. Let S be a set, and let $F(S)$ be the free group on S. Call a reduced word $s_0s_1 \ldots s_{t-1} \in F(S)$ (with $s_i \in S \cup \bar{S}$) *cyclically reduced* if $s_0 \ne \bar{s}_{t-1}$ or $t = 0$. Prove that every conjugacy class in $F(S)$ contains a cyclically reduced word. Which conjugacy classes contain a *unique* cyclically reduced word? And which contain infinitely many?

5.17. Let A and B be groups, and let $w \in A \circ B$. Prove: w has finite order if and only if w is conjugate to an element of finite order of A or B (viewed as subgroups of $A \circ B$).

5.18. For a group H, write $H^{\mathrm{ab}} = H/[H,H]$.

(a) Let H and G be groups. Exhibit a bijection between $\mathrm{Hom}(H^{\mathrm{ab}}, G)$ and $\{f \in \mathrm{Hom}(H,G) : fH$ is an abelian subgroup of $G\}$. Formulate what it means for your isomorphism to be functorial in G, and prove that it is. Is it also functorial in H?

(b) Prove that for any two groups A and B one has $(A \circ B)^{\mathrm{ab}} \cong A^{\mathrm{ab}} \times B^{\mathrm{ab}}$. Do this by describing the homomorphisms from these two groups to a test group G.

5.19. Let S be a set, and write $\mathbf{Z}^{(S)}$ for the group of all functions $f\colon S \to \mathbf{Z}$ with the property that the *support* $\{s \in S : f(s) \ne 0\}$ of f is a *finite* subset of S; the additively written group operation on $\mathbf{Z}^{(S)}$ is defined by $(f_1+f_2)(s) = f_1(s)+f_2(s)$, for $f_1, f_2 \in \mathbf{Z}^{(S)}$, $s \in S$. We call $\mathbf{Z}^{(S)}$ the *free abelian group on S*.

(a) Prove: $\mathbf{Z}^{(S)} \cong F(S)^{\mathrm{ab}}$, the notation being as in Exercise 5.18. Do this by describing the homomorphisms from these two groups to a test group G.

(b) Suppose T is another set. Prove that the following three properties are equivalent: (i) $F(S)$ and $F(T)$ are isomorphic as groups; (ii) $\mathbf{Z}^{(S)}$ and $\mathbf{Z}^{(T)}$ are isomorphic as groups; (iii) $\#S = \#T$. If you cannot do this, restrict to the case S is finite.

5.20. Let S be a well-ordered set, i. e. a totally ordered set such that every non-empty subset of S contains a smallest element. Let $\mathbf{Z}^{(S)}$ be as in Exercise 5.19. For $f \in \mathbf{Z}^{(S)}$, $f \ne 0$, we write $\deg f = \max\{s \in S : f(s) \ne 0\}$ (this is well-defined, since $\{s \in S : f(s) \ne 0\}$ is finite).

Prove: if $H \subset \mathbf{Z}^{(S)}$ is a subgroup, and $T = \{\deg f : f \in H, f \neq 0\}$, then one has $H \cong \mathbf{Z}^{(T)}$.

5.21. An abelian group is called *free* if it is isomorphic to $\mathbf{Z}^{(S)}$ for some set S. Prove: any subgroup of a free abelian group is a free abelian group.

Note. Likewise, a (not necessarily abelian) group is called *free* if it is isomorphic to $F(S)$ for some set S. (One should be careful with this terminology: $\mathbf{Z} \times \mathbf{Z}$ is free when considered as an abelian group, but it is not free when considered as a group.) It is again true that subgroups of free groups are free groups, but the proof is much more involved.

5.22. Let H be a subgroup of finite index of a finitely generated group. Prove that H is finitely generated.

5.23. (a) Let G be the set of those pairs $\big((n_i)_{i\in\mathbf{Z}}, (k_j)_{j\in\mathbf{Z},j>0}\big)$ of infinite vectors of integers n_i and k_j for which the sets $\{i : n_i \neq 0\}$ and $\{j : k_j \neq 0\}$ are finite. Prove that the operation $*$ on G defined by

$$\big((n_i)_i, (k_j)_j\big) * \big((m_i)_i, (l_j)_j\big) = \big((n_i + m_i)_i, (k_j + l_j + \textstyle\sum_{i\in\mathbf{Z}} n_i m_{i+j})_j\big)$$

makes G into a group. Prove also that the map $\tau: G \to G$ sending $\big((n_i)_{i\in\mathbf{Z}}, (k_j)_{j\in\mathbf{Z},j>0}\big)$ to $\big((n_{i-1})_{i\in\mathbf{Z}}, (k_j)_{j\in\mathbf{Z},j>0}\big)$ is an automorphism of G.

(b) Let $\langle\tau\rangle$ be the subgroup of $\operatorname{Aut} G$ generated by τ. Prove that the semidirect product $F = G \rtimes \langle\tau\rangle$ can be generated, as a group, by two elements, and that the center $Z(F)$ of F is not a finitely generated group.

5.24. (a) Prove: if a group F is finitely generated, and its center $Z(F)$ is not finitely generated, then $F/Z(F)$ is finitely generated but not finitely presented.

(b) Prove that the group $\langle x, y : xy^jxy^{-j} = y^jxy^{-j}x \ (j = 1, 2, 3, \ldots)\rangle$ is finitely generated but not finitely presented.

6. Tensor products

In this section, we write all groups additively, although we do not assume that the groups are abelian. We do this because we shall discover that, for the purposes of forming tensor products, abelian groups are in a sense the only interesting ones.

Suppose A, B, and G are groups. A map $f: A \times B \to G$ is called *bilinear* if for all a, $a' \in A$ and b, $b' \in B$ one has

$$f(a, b + b') = f(a, b) + f(a, b') \quad \text{and} \quad f(a + a', b) = f(a, b) + f(a', b).$$

In other words, f is bilinear if and only if for each $a \in A$ the map $b \mapsto f(a, b)$ is a homomorphism $B \to G$ and for each $b \in B$ the map $a \mapsto f(a, b)$ is a homomorphism $A \to G$. In particular, for every $a \in A$ and $b \in B$ one has $f(a, 0) = 0$ and $f(0, b) = 0$. (Do not confuse bilinear maps with group homomorphisms $A \times B \to G$. If a bilinear map is a group homomorphism, then for all $a \in A$ and $b \in B$ one has $f(a, b) = f(a, 0) + f(0, b) = 0 + 0 = 0$, so f is the zero map.) We shall write $\mathrm{Bil}(A \times B, G)$ for the set of all bilinear maps $A \times B \to G$.

Examples. (i) Let R be a ring, and R^+ its additive group. Then the multiplication map $R^+ \times R^+ \to R^+$, $(r, s) \mapsto rs$, is bilinear, by the two distributive laws.

(ii) Let R be a ring, and M an R-module. Then the scalar multiplication map $R^+ \times M \to M$, $(r, m) \mapsto rm$, is bilinear, by two of the module axioms.

(iii) Let us describe the bilinear maps $f: \mathbf{Z} \times B \to G$. To know f, it suffices to know the function $g: B \to G$ defined by $g(b) = f(1, b)$. Namely, if b is fixed then the map $n \to f(n, b)$ is a group homomorphism $\mathbf{Z} \to G$, so one has $f(n, b) = n \cdot f(1, b) = n \cdot g(b)$. Thus, the question is for which functions $g: B \to G$ the expression $n \cdot g(b)$ is bilinear in n and b. We know already that g must be a group homomorphism. Also, $2 \cdot g$ is a group homomorphism, so for all b, $b' \in B$ one has $2 \cdot g(b + b') = 2 \cdot g(b) + 2 \cdot g(b')$; that is, $g(b) + g(b') + g(b) + g(b') = g(b) + g(b) + g(b') + g(b')$, so $g(b') + g(b) = g(b) + g(b')$. Hence the image of g is an abelian subgroup of G, or, equivalently, one has $[B, B] \subset \ker g$. Conversely, if $g: B \to G$ is a group homomorphism with an abelian image, then one readily checks that the expression $n \cdot g(b)$ is bilinear in n and b. We conclude that there are bijections

$$\mathrm{Bil}(\mathbf{Z} \times B, G) \cong \{g \in \mathrm{Hom}(B, G) : gB \text{ is abelian}\} \cong \mathrm{Hom}(B/[B, B], G).$$

One readily checks that these bijections are functorial in G.

(iv) Clearly, one has $\mathrm{Bil}(0 \times B, G) \cong \mathrm{Hom}(0, G)$ for every group G.

(v) For any bilinear map $f: (\mathbf{Z}/2\mathbf{Z}) \times B \to G$ we have $f(0, b) = 0$ and $f(1, b) = g(b)$ for some group homomorphism $g: B \to G$. One must have $g(b) + g(b) = f(1+1, b) = f(0, b) = 0$, so each element of $g(B)$ is its own negative; equivalently, $\ker g$ contains the subgroup $2B = \langle 2b : b \in B \rangle$ of B. The subgroup $2B$ is normal in B, because it is characteristic.

Also, it contains $[B, B]$ because $b + b' - b - b' = 2b + 2(-b + b') + 2(-b')$. Conversely, if $g: B \to G$ is a group homomorphism with $2B \subset \ker g$, then the image of g is abelian because $[B, B] \subset \ker g$, and one deduces that the map $(\mathbf{Z}/2\mathbf{Z}) \times B \to G$ sending $(0, b)$ to 0 and $(1, b)$ to $g(b)$ is bilinear. Altogether one has a bijection

$$\mathrm{Bil}((\mathbf{Z}/2\mathbf{Z}) \times B, G) \cong \mathrm{Hom}(B/2B, G)$$

that is functorial in G.

For general groups A and B, the tensor product $A \otimes B$ that we shall define and study in this section is characterized by the existence of bijections

$$\mathrm{Bil}(A \times B, G) \cong \mathrm{Hom}(A \otimes B, G)$$

that are functorial in G. Thus, we can express the results of the last three examples by saying that

$$\mathbf{Z} \otimes B \cong B/[B, B], \quad 0 \otimes B = 0, \quad (\mathbf{Z}/2\mathbf{Z}) \otimes B \cong B/2B.$$

Lemma 1. *Let A and B be groups, put $A^{\mathrm{ab}} = A/[A, A]$ and $B^{\mathrm{ab}} = B/[B, B]$, and let $c: A \times B \to A^{\mathrm{ab}} \times B^{\mathrm{ab}}$ be the canonical map. Then for each group G there is a bijection*

$$\mathrm{Bil}(A^{\mathrm{ab}} \times B^{\mathrm{ab}}, G) \xrightarrow{\sim} \mathrm{Bil}(A \times B, G)$$

that sends g to $g \cdot c$, and that is functorial in G. Also, for each group G and each bilinear map $f: A \times B \to G$ the subgroup of G generated by the image of f is abelian.

Proof. One readily verifies that $g \cdot c: A \times B \to G$ is bilinear for each $g \in \mathrm{Bil}(A^{\mathrm{ab}} \times B^{\mathrm{ab}}, G)$, so the map $\mathrm{Bil}(A^{\mathrm{ab}} \times B^{\mathrm{ab}}, G) \to \mathrm{Bil}(A \times B, G)$ is well-defined. It is also injective, because c is surjective. Functoriality is easy to verify. To prove that it is surjective, we first prove the last statement of the lemma. Let $f \in \mathrm{Bil}(A \times B, G)$, and a, $a' \in A$, b, $b' \in B$. One has

$$\begin{aligned} f(a,b) + f(a,b') + f(a',b) + f(a',b') &= f(a, b+b') + f(a', b+b') = f(a+a', b+b') \\ &= f(a+a', b) + f(a+a', b') = f(a,b) + f(a',b) + f(a,b') + f(a',b') \end{aligned}$$

so $f(a,b') + f(a',b) = f(a',b) + f(a,b')$. Hence any two elements of the image of f commute, so $\mathrm{im}\, f$ generates an abelian subgroup of G. It follows that for fixed $b \in B$ the kernel of the group homomorphism $A \to G$, $a \mapsto f(a, b)$, contains $[A, A]$, so $f(a, b)$ depends only on the coset $a + [A, A]$ of a. Likewise, for fixed $a \in A$, the element $f(a, b)$ of G depends only on the coset $b + [B, B]$. This implies that $f = g \cdot c$ for some map $g: A^{\mathrm{ab}} \times B^{\mathrm{ab}} \to G$. Using that c is surjective one verifies that g is bilinear. This proves Lemma 1.

Theorem 2. *Let A and B be groups. Then there exists a group $A \otimes B$ and a bilinear map $\otimes\colon A \times B \to A \otimes B$ such that for any group G the map $\otimes^*\colon \mathrm{Hom}(A \otimes B, G) \to \mathrm{Bil}(A \times B, G)$ defined by $h \mapsto h \circ \otimes$ is bijective. Also, this property determines the pair $(A \otimes B, \otimes)$ up to a unique isomorphism.*

The last statement in Theorem 2 means the following: if C is any group with a bilinear map $c\colon A \times B \to C$ such that for every group G the similarly defined map $c^*\colon \mathrm{Hom}(C, G) \to \mathrm{Bil}(A \times B, G)$ is bijective, then there is a unique group isomorphism $h\colon A \otimes B \to C$ such that $c = h \circ \otimes$.

The group $A \otimes B$ is called the *tensor product* of A and B, and for $a \in A$ and $b \in B$ the image of (a, b) under $\otimes$ in $A \otimes B$ is written $a \otimes b$.

Proof of Theorem 2. Define $A \otimes B = \langle S\colon R\rangle$ where S is the set $A \times B$ and R is given by $R = \{(a, b + b') - (a, b) - (a, b') : a \in A,\ b,\ b' \in B\} \cup \{(a + a', b) - (a, b) - (a', b) : a,\ a' \in A,\ b \in B\}$. Also, define $a \otimes b$ to be the image of $(a, b) \in S$ in $\langle S\colon R\rangle$. From Proposition 4 of Section 5 we know that $\otimes^*$ is a functorial bijection from the set $\mathrm{Hom}(A \otimes B, G) = \mathrm{Hom}(\langle S\colon R\rangle, G)$ to the set of maps $S = A \times B \to G$ satisfying R; the latter set is just the set $\mathrm{Bil}(A \times B, G)$ of bilinear maps $A \times B \to G$. Applying this to $G = A \otimes B$, we see that the identity homomorphism $1_{A \otimes B} \in \mathrm{Hom}(A \otimes B, A \otimes B)$ gives rise to the map $\otimes \in \mathrm{Bil}(A \times B, A \otimes B)$; that is, the latter map is bilinear. The uniqueness statement follows from Theorem 3 of Section 5. This proves Theorem 2.

Theorem 3. *Let A and B be groups. Then the image of $\otimes\colon A \times B \to A \otimes B$ generates $A \otimes B$, the group $A \otimes B$ is abelian, and there is an isomorphism $A^{\mathrm{ab}} \otimes B^{\mathrm{ab}} \xrightarrow{\sim} A \otimes B$.*

Proof. Since the set $S = A \times B$ considered in the previous proof generates $F(S)$, and the map $F(S) \to A \otimes B$ is surjective, the group $A \otimes B$ is generated by the image of $A \times B$. This proves the first assertion of Theorem 3, and the second follows from Lemma 1. The last assertion of Theorem 3 follows from Lemma 1 combined with Theorem 3 of Section 5. This proves Theorem 3.

Theorem 3 and the rule $-(a \otimes b) = (-a) \otimes b$ imply that every element of $A \otimes B$ can be written as a finite sum $a_1 \otimes b_1 + \ldots + a_t \otimes b_t$ with $a_i \in A$, $b_i \in B$; however, this representation is *not* unique. Also, it is *not* generally true that every element of $A \otimes B$ is of the form $a \otimes b$, with $a \in A$, $b \in B$ (cf. Exercise 6.8(b)).

The theorem shows that without true loss of generality one may restrict the formation of tensor product to the case of *abelian* groups. Our earlier example $\mathbf{Z} \otimes B \cong B^{\mathrm{ab}}$ simply reads $\mathbf{Z} \otimes B \cong B$ when B is abelian.

Example. As an example of an explicit computation of a tensor product, we next show that *there is an isomorphism $\mathbf{Q} \otimes \mathbf{Q} \xrightarrow{\sim} \mathbf{Q}$ that maps $a \otimes b$ to ab, for a, $b \in \mathbf{Q}$.*

The map $\mathbf{Q} \times \mathbf{Q} \to \mathbf{Q}$ sending (a, b) to ab is bilinear, so by the defining property of tensor products there is a unique group homomorphism $\varphi: \mathbf{Q} \otimes \mathbf{Q} \to \mathbf{Q}$ with $\varphi(a \otimes b) = ab$. Also, since $\otimes$ is bilinear, the map $\psi: \mathbf{Q} \to \mathbf{Q} \otimes \mathbf{Q}$ defined by $\psi(b) = 1 \otimes b$ is a group homomorphism. We claim that φ and ψ are a pair of inverse mappings.

It is clear that $\varphi\psi = 1_{\mathbf{Q}}$. In the proof of $\psi\varphi = 1_{\mathbf{Q}\otimes\mathbf{Q}}$, we shall use the general rule

$$(na) \otimes b = n(a \otimes b) = a \otimes (nb) \qquad \text{in } A \otimes B,$$

which is valid for any two abelian groups A, B, any $n \in \mathbf{Z}$, and any $a \in A$, $b \in B$. To prove this rule, one observes that both sums are equal to $a \otimes b + a \otimes b + \cdots + a \otimes b$ (with n terms). A formal proof would use induction on n when n is non-negative, and deduce the case of negative n from the case of positive n.

Now consider an arbitrary element $a_1 \otimes b_1 + \cdots + a_t \otimes b_t$ of $\mathbf{Q} \otimes \mathbf{Q}$, with a_i, $b_i \in \mathbf{Q}$. Let $n \in \mathbf{Z}$, $n > 0$, be such that $a_i \in \frac{1}{n}\mathbf{Z}$ for each i. With $na_i = c_i \in \mathbf{Z}$, one has $a_i \otimes b_i = a_i \otimes (n\frac{b_i}{n}) = (na_i) \otimes \frac{b_i}{n} = c_i \otimes \frac{b_i}{n} = (c_i 1) \otimes \frac{b_i}{n} = 1 \otimes \frac{c_i b_i}{n}$ for each i, and therefore $a_1 \otimes b_1 + \cdots + a_t \otimes b_t = 1 \otimes (\sum_i \frac{c_i b_i}{n})$. So every element of $\mathbf{Q} \otimes \mathbf{Q}$ is of the form $1 \otimes r$ with $r \in \mathbf{Q}$. This clearly implies that $\psi\varphi = 1_{\mathbf{Q}\otimes\mathbf{Q}}$. This proves the claim and completes the proof that $\mathbf{Q} \otimes \mathbf{Q} \cong \mathbf{Q}$.

The computation that we just gave is not typical for the computation of tensor products. Usually it is more efficient to use general properties of tensor products, as listed in Theorems 4 and 6 below, in order to reduce the computation of a tensor product $A \otimes B$ to a case that one knows already, for example the case in which $A = \mathbf{Z}$.

The *direct sum* $\bigoplus_{i \in I} B_i$ of a family $(B_i)_{i \in I}$ of abelian groups is defined to be the subgroup $\{(b_i)_{i \in I} : b_i \in B_i, \#\{i \in I: b_i \neq 0\} < \infty\}$ of the product group $\prod_{i \in I} B_i = \{(b_i)_{i \in I}: b_i \in B_i$ for all $i \in I\}$; note that the sum is different from the product if I is infinite and infinitely many B_i's are non-zero.

Theorem 4. *One has*

$$A \otimes B \cong B \otimes A,$$
$$(A \otimes B) \otimes C \cong A \otimes (B \otimes C),$$
$$A \otimes \left(\bigoplus_{i \in I} B_i\right) \cong \bigoplus_{i \in I} (A \otimes B_i)$$

for any three abelian groups A, B, C and any family $(B_i)_{i \in I}$ of abelian groups.

Proof. We deduce all three properties from Theorem 3 of Section 5. For the first property, it suffices to observe that there is an obvious bijection $\mathrm{Bil}(A \times B, G) \to \mathrm{Bil}(B \times A, G)$, functorial in G, that maps f to g if $g(b, a) = f(a, b)$. It leads to an isomorphism $A \otimes B \xrightarrow{\sim} B \otimes A$ that maps $a \otimes b$ to $b \otimes a$. For the second property, it is convenient to define

$\mathrm{Tril}(A \times B \times C, G)$ as the set of *trilinear* maps $f\colon A \times B \times C \to G$, that is, maps satisfying

$$\begin{aligned} f(a + a', b, c) &= f(a, b, c) + f(a', b, c) \\ f(a, b + b', c) &= f(a, b, c) + f(a, b', c) \\ f(a, b, c + c') &= f(a, b, c) + f(a, b, c') \end{aligned}$$

for all $a,\ a' \in A,\ b,\ b' \in B$, and $c,\ c' \in C$. One now readily verifies that there are bijections $\mathrm{Hom}\big((A \otimes B) \otimes C, G\big) \cong \mathrm{Bil}\big((A \otimes B) \times C, G\big) \cong \mathrm{Tril}(A \times B \times C, G) \cong \mathrm{Bil}\big(A \times (B \otimes C), G\big) \cong \mathrm{Hom}\big(A \otimes (B \otimes C), G\big)$ that are functorial in G, and that the resulting isomorphism $(A \otimes B) \otimes C \xrightarrow{\sim} A \otimes (B \otimes C)$ maps $(a \otimes b) \otimes c$ to $a \otimes (b \otimes c)$. The last property follows from the functorial bijections

$$\mathrm{Hom}\big(A \otimes (\textstyle\bigoplus_{i \in I} B_i), G\big) \cong \mathrm{Bil}\big(A \times (\textstyle\bigoplus_{i \in I} B_i), G\big) \cong$$
$$\prod_{i \in I} \mathrm{Bil}(A \times B_i, G) \cong \prod_{i \in I} \mathrm{Hom}(A \otimes B_i, G) \cong \mathrm{Hom}\big(\textstyle\bigoplus_{i \in I}(A \otimes B_i), G\big).$$

It induces an isomorphism $A \otimes \big(\bigoplus_{i \in I} B_i\big) \xrightarrow{\sim} \bigoplus_{i \in I}(A \otimes B_i)$ that sends $a \otimes (b_i)_{i \in I}$ to $(a \otimes b_i)_{i \in I}$. This proves Theorem 4.

Next we consider tensor products of *group homomorphisms.*

Theorem 5. *Let $f\colon A \to C$ and $g\colon B \to D$ be group homomorphisms. Then there exists a unique group homomorphism $f \otimes g\colon A \otimes B \to C \otimes D$ that maps $a \otimes b$ to $f(a) \otimes g(b) \in C \otimes D$ for all $a \in A$ and $b \in B$.*

Proof. From the fact that f and g are group homomorphisms and $\otimes\colon C \times D \to C \otimes D$ is bilinear one deduces that that map $A \times B \to C \otimes D$ sending (a, b) to $f(a) \otimes g(b)$ is bilinear. Hence, by the characterizing property of $A \otimes B$, there is a unique homomorphism from $A \otimes B$ to $C \otimes D$ that maps $a \otimes b$ to $f(a) \otimes g(b) \in C \otimes D$ for all $a \in A$ and $b \in B$. This proves Theorem 5.

One checks in a straightforward way the following properties of tensor products of maps:

$$\begin{gathered} 1_A \otimes (g \circ h) = (1_A \otimes g) \circ (1_A \otimes h), \\ (e \circ f) \otimes 1_B = (e \otimes 1_B) \circ (f \otimes 1_B), \\ 1_A \otimes 1_B = 1_{A \otimes B}, \\ f \otimes (g + g') = (f \otimes g) + (f \otimes g'), \\ (f + f') \otimes g = (f \otimes g) + (f' \otimes g), \end{gathered}$$

which are valid whenever they are meaningful; for example, for the third property it is supposed that $f\colon A \to A',\ g,\ g'\colon B \to C$ are homomorphisms between abelian groups, and the maps asserted to be equal go from $A \otimes B$ to $A' \otimes C$. The third property may be phrased by saying that tensoring with f is a *group homomorphism* $\mathrm{Hom}(B, C) \to \mathrm{Hom}(A \otimes B, A' \otimes C)$; in particular, it implies that $f \otimes 0 = 0$.

As an application, we give a different proof of the isomorphism $A \otimes (B \oplus C) \cong (A \otimes B) \oplus (A \otimes C)$ from Theorem 4. The proof is 'categorical' in the sense of Section 7, which means that it just manipulates with 'arrows'. It uses only of the properties of the tensor products of maps listed above, and consequently it applies to any construction that shares these properties with the tensor product. The basis of the proof is formed by an 'arrow-theoretic' characterization of the direct sum.

Let B, C be abelian groups. Define the functions

$$i_1\colon B \to B \oplus C, \quad \pi_1\colon B \oplus C \to B, \quad i_2\colon C \to B \oplus C, \quad \pi_2\colon B \oplus C \to C$$

by $i_1(b) = (b, 0)$, $\pi_1(b, c) = b$, $i_2(c) = (0, c)$, and $\pi_2(b, c) = c$ for $b \in B$, $c \in C$. These are group homomorphisms, and they satisfy

$$\pi_1 i_1 = 1_B, \quad \pi_2 i_2 = 1_C, \quad \pi_1 i_2 = 0, \quad \pi_2 i_1 = 0, \quad i_1\pi_1 + i_2\pi_2 = 1_{B\oplus C},$$

where the addition in the last formula takes place in the additive group $\mathrm{Hom}(B\oplus C, B\oplus C)$. These properties *characterize* $B \oplus C$. That is, if D is an abelian group, and there are group homomorphisms $i_1'\colon B \to D$, $\pi_1'\colon D \to B$, $i_2'\colon C \to D$, $\pi_2'\colon D \to C$ that have the corresponding properties, then there is a isomorphism $D \cong B \oplus C$; more precisely, the maps $i_1\pi_1' + i_2\pi_2'\colon D \to B\oplus C$ and $i_1'\pi_1 + i_2'\pi_2\colon B\oplus C \to D$ are inverse group isomorphisms. The verification is left to the reader.

Now let A be an abelian group. To deduce that $A \otimes (B \oplus C) \cong (A \otimes B) \oplus (A \otimes C)$, we define the maps

$$\begin{aligned} i_1'&\colon A \otimes B \to A \otimes (B \oplus C), \qquad & \pi_1'&\colon A \otimes (B \oplus C) \to A \otimes B,\\ i_2'&\colon A \otimes C \to A \otimes (B \oplus C), \qquad & \pi_2'&\colon A \otimes (B \oplus C) \to A \otimes C \end{aligned}$$

by $i_1' = 1_A \otimes i_1$, $\pi_1' = 1_A \otimes \pi_1$, $i_2' = 1_A \otimes i_2$, and $\pi_2' = 1_A \otimes \pi_2$. Next we apply the above characterization of the direct sum, with $A\otimes(B\oplus C)$, $A\otimes B$, and $A\otimes C$ in the roles of D, B, and C, respectively. Then we find that in order to prove that $A\otimes(B\oplus C) \cong (A\otimes B)\oplus(A\otimes C)$ it suffices to verify that

$$\begin{gathered} (1_A \otimes \pi_1)(1_A \otimes i_1) = 1_{A\otimes B}, \quad (1_A \otimes \pi_2)(1_A \otimes i_2) = 1_{A\otimes C},\\ (1_A \otimes \pi_1)(1_A \otimes i_2) = 0, \quad (1_A \otimes \pi_2)(1_A \otimes i_1) = 0,\\ (1_A \otimes i_1)(1_A \otimes \pi_1) + (1_A \otimes i_2)(1_A \otimes \pi_2) = 1_{A\otimes(B\oplus C)}. \end{gathered}$$

The first follows from $(1_A \otimes \pi_1)(1_A \otimes i_1) = 1_A \otimes (\pi_1 i_1) = 1_A \otimes 1_B = 1_{A\otimes B}$, and the others are proved similarly.

Other important properties of the tensor product of maps are

$$(e \otimes f) \circ (g \otimes h) = (e \circ g) \otimes (f \circ h) \qquad (f \otimes g) \otimes h = f \otimes (g \otimes h),$$

which are again valid when they are meaningful; for the last property, it is assumed that $(A \otimes B) \otimes C$ is identified with $A \otimes (B \otimes C)$, as in Theorem 4.

We now come to one of the most important tools used in computing tensor products.

Theorem 6. *Let A be a group, and let $B \xrightarrow{f} C \xrightarrow{g} D \to 0$ be an exact sequence of abelian groups. Then the sequence $A \otimes B \xrightarrow{1 \otimes f} A \otimes C \xrightarrow{1 \otimes g} A \otimes D \to 0$ (with $1 = 1_A$) is exact.*

The property formulated in Theorem 6 is expressed by saying that tensoring is *right exact*.

Proof of Theorem 6. Exactness of the first sequence is equivalent to g inducing an isomorphism $D \cong \operatorname{cok} f = C/fB$. Hence it suffices to show that tensoring with A preserves cokernels, i.e. $A \otimes \operatorname{cok} f \cong \operatorname{cok}(1_A \otimes f)$. Let G be any group. Then there are functorial bijections

$$\begin{aligned} \operatorname{Hom}(A \otimes \operatorname{cok} f, G) &\cong \operatorname{Bil}(A \times (C/fB), G) \cong \{h \in \operatorname{Bil}(A \times C, G) : h \circ (1_A \times f) = 0\} \\ &\cong \operatorname{Hom}((A \otimes C)/\operatorname{im}(1_A \otimes f), G) = \operatorname{Hom}(\operatorname{cok}(1_A \otimes f), G). \end{aligned}$$

It follows that $A \otimes \operatorname{cok} f \cong \operatorname{cok}(1_A \otimes f)$. This proves Theorem 6.

Tensoring is *not* left exact, i.e., if $0 \to B \to C \to D$ is exact then in general $0 \to A \otimes B \to A \otimes C \to A \otimes D$ need not be exact. To give an example, consider the exact sequence $0 \to \mathbf{Z} \xrightarrow{2} \mathbf{Z} \to \mathbf{Z}/2\mathbf{Z} \to 0$, the map $\mathbf{Z} \to \mathbf{Z}$ being multiplication by 2, and let A be any abelian group. The map $1_A \otimes 2\colon A \cong A \otimes \mathbf{Z} \to A \otimes \mathbf{Z} \cong A$ equals $1_A \otimes (1_{\mathbf{Z}} + 1_{\mathbf{Z}}) = 1_A + 1_A$, so it is multiplication by 2. Thus, tensoring with A we obtain the sequence $0 \to A \xrightarrow{2} A \to A \otimes (\mathbf{Z}/2\mathbf{Z}) \to 0$, which is *not* exact if A has an element of order 2. The sequence *is* exact on the right, which confirms the earlier result that $A \otimes (\mathbf{Z}/2\mathbf{Z}) \cong A/2A$.

Thus, in general tensoring does not preserve injectivity of maps. Since an injective group homomorphism $B \to C$ may be viewed as an identificaton of B with a subgroup of C, we also see that *if B is a subgroup of C, then $A \otimes B$ need* not *be a subgroup of $A \otimes C$.*

Example. Let n and m be positive integers. Then one has

$$(\mathbf{Z}/n\mathbf{Z}) \otimes (\mathbf{Z}/m\mathbf{Z}) \cong \mathbf{Z}/\gcd(n,m)\mathbf{Z}.$$

To prove this, tensor the exact sequence $\mathbf{Z} \xrightarrow{m} \mathbf{Z} \to \mathbf{Z}/m\mathbf{Z} \to 0$ with $\mathbf{Z}/n\mathbf{Z}$. By Theorem 6 and the isomorphism $(\mathbf{Z}/n\mathbf{Z}) \otimes \mathbf{Z} \cong \mathbf{Z}/n\mathbf{Z}$ one obtains an exact sequence

$$\mathbf{Z}/n\mathbf{Z} \xrightarrow{m} \mathbf{Z}/n\mathbf{Z} \to (\mathbf{Z}/n\mathbf{Z}) \otimes (\mathbf{Z}/m\mathbf{Z}) \to 0,$$

and the result now follows from $m(\mathbf{Z}/n\mathbf{Z}) = (m\mathbf{Z} + n\mathbf{Z})/n\mathbf{Z} = \gcd(n,m)\mathbf{Z}/n\mathbf{Z}$.

6. Exercises

6.1. Let n be an integer, A an abelian group, and $n_A: A \to A$ the map $a \mapsto na$ from Exercise 4.4. Prove: $(\mathbf{Z}/n\mathbf{Z}) \otimes A \cong \operatorname{cok} n_A$. Formulate what it means for your isomorphism to be functorial in A, and prove that it is.

6.2. Describe the group $A \otimes B$ when each of A and B is one of the following: (a) finite cyclic; (b) infinite cyclic; (c) the Klein four group; (d) the additive group of a vector space over $\mathbf{Q}$; and (e) $\mathbf{Q}/\mathbf{Z}$. (Be sure to cover all combinations.)

6.3. Is the tensor product of two finitely generated abelian groups finitely generated? Is the tensor product of two finite abelian groups finite? Give in each case a proof or a counterexample.

6.4. Suppose that A and B are non-zero finitely generated abelian groups. Prove: $A \otimes B = 0$ if and only if A and B are finite with $\gcd(\#A, \#B) = 1$.

6.5. (a) An abelian group A is called *divisible* if for each positive integer n the map $n_A: A \to A$ from Exercise 4.4 is surjective. Prove: if A is divisible, then $A \otimes B$ is divisible for each abelian group B.

(b) An abelian group A is called *torsion* if each $a \in A$ is of finite order. Prove: if A is torsion, then $A \otimes B$ is torsion for each abelian group B.

(c) Give an example of a non-zero divisible torsion abelian group. Is there a finite one?

6.6. Let A and B be abelian groups, with A divisible and B torsion. Prove: $A \otimes B = 0$.

6.7. The *rank* of a free abelian group A is defined to be $\#S$ when $A \cong \mathbf{Z}^{(S)}$ (see Exercise 5.19); by Exercise 5.19(b), this is well-defined.

Suppose that A and B are free abelian groups. Prove that $A \otimes B$ is free as well, and express the rank of $A \otimes B$ in terms of the rank of A and the rank of B. (If you are uncomfortable with cardinal numbers, restrict to the case of finite rank.)

6.8. (a) Give an example of groups A, B, C and a bilinear map $f: A \times B \to C$ such that the image of f is not a subgroup of C.

(b) Give an example of groups A, B such that not every element of $A \otimes B$ is of the form $a \otimes b$, with $a \in A$, $b \in B$.

6.9. Let G be a group. The *second exterior power* $G \wedge G$ of G is defined to be the group $(G \otimes G)/N$, where N is the subgroup of $G \otimes G$ generated by $\{x \otimes x : x \in G\}$ (it is normal since $G \otimes G$ is abelian). For $x, y \in G$ one defines $x \wedge y \in G \wedge G$ to be the image of $x \otimes y$ in $G \wedge G$.

(a) Prove that $x \wedge y = -(y \wedge x)$, for all $x, y \in G$.

(b) Suppose that $G^{\mathrm{ab}} = G/[G, G]$ is cyclic. Prove that $G \wedge G$ is trivial.

(c) What is the order of $V_4 \wedge V_4$?

6.10. Let G be a group, and for g, $h \in G$ write $[g,h] = ghg^{-1}h^{-1}$ (in multiplicative notation; in additive notation it would be $g + h - g - h$).

(a) Prove: the map $G \times G \to [G,G]$ sending (g,h) to $[g,h]$ is bilinear if and only if $[G,G]$ is contained in the center $Z(G)$ of G.

(b) Let G be a group with $[G,G] \subset Z(G)$. Show that there is a surjective group homomorphism $G \wedge G \to [G,G]$ (see Exercise 6.9 for the definition of $G \wedge G$).

(c) Exhibit a finite non-abelian group G satisfying $[G,G] \subset Z(G)$ and $G \wedge G \cong [G,G]$.

6.11. Let G be a multiplicatively written group with $[G,G] \subset Z(G)$.

(a) Suppose that $\sqrt{}\colon [G,G] \to Z(G)$ is a group homomorphism with the property $\left(\sqrt{c}\right)^2 = c$ for every $c \in [G,G]$. Define an operation $*$ on G by $a * b = ab\sqrt{[b,a]}$, for a, $b \in G$. Prove that $*$ makes G into an abelian group, and that for a, $b \in G$ one has $a * b = ab$ if and only if $ab = ba$.

(b) Suppose that $[G,G]$ is finite. Prove that a map $\sqrt{}$ as in (a) exists if and only if $[G,G]$ has odd order, that in this case the map $\sqrt{}$ is unique, and that every subgroup of G remains a subgroup under $*$.

(c) Exhibit a finite non-abelian group to which the construction above applies.

6.12. (a) Let A be a finite abelian group, and let $f\colon A \wedge A \to \mathbf{Q}/\mathbf{Z}$ be a group homomorphism with the property that for every non-zero $a \in A$ there exists $b \in A$ with $f(a \wedge b) \neq 0$. Prove that there is a finite abelian group B and an isomorphism $\psi\colon B \oplus \hat{B} \to A$ (where $\hat{B} = \mathrm{Hom}(B, \mathbf{Q}/\mathbf{Z})$) such that for any b_1, $b_2 \in B$ and g_1, $g_2 \in \hat{B}$ one has $f\big(\psi(b_1, g_1) \wedge \psi(b_2, g_2)\big) = g_1(b_2) - g_2(b_1)$. (This takes some work.)

(b) Let G be a finite group for which $[G,G]$ is cyclic and contained in $Z(G)$. Prove that $G/Z(G)$ has square order m^2, and that $m = \min\{(G : H) : H$ is an abelian subgroup of $G\}$.

6.13. Let I be a set, let B_i an abelian group for each $i \in I$, and let $\mathcal{F}_{ij}$ be a set of group homomorphisms $B_i \to B_j$, for every $(i,j) \in I \times I$. A *right limit* (or *colimit*, or *direct limit*) of the system $\big((B_i)_{i \in I}, (\mathcal{F}_{ij})_{(i,j) \in I \times I}\big)$ is an abelian group B with the property that there is, for each abelian group C, a bijection (functorial in C) between the set $\mathrm{Hom}(B, C)$ and the set of those systems $(g_i)_{i \in I}$ of group homomorphisms $g_i\colon B_i \to C$ that have the property $g_j f = g_i$ for any i, $j \in I$ and any $f \in \mathcal{F}_{ij}$.

(a) Prove that any system $\big((B_i)_{i \in I}, (\mathcal{F}_{ij})_{(i,j) \in I \times I}\big)$ possesses a right limit, and that it is uniquely determined up to a canonical isomorphism. The right limit is denoted by

$$\varinjlim \big((B_i)_{i \in I}, (\mathcal{F}_{ij})_{(i,j) \in I \times I}\big),$$

or simply by

$$\varinjlim_{i \in I} B_i$$

if the sets $\mathcal{F}_{ij}$ are clear from the context.

(b) Prove that the notion of right limits generalizes the notion of (possibly infinite) direct sums, and that it also generalizes the notion of cokernels.

6.14. Prove that "tensoring commutes with right limits" in the following sense. Let A be an abelian group, and let $\big((B_i)_{i\in I}, (\mathcal{F}_{ij})_{(i,j)\in I\times I}\big)$ be a system as in Exercise 6.13. Then there is an isomorphism

$$A \otimes \big(\varinjlim_{i\in I} B_i\big) \cong \varinjlim_{i\in I}(A \otimes B_i),$$

the limit on the right being taken with respect to the collections $1_A \otimes \mathcal{F}_{ij} = \{1_A \otimes f : f \in \mathcal{F}_{ij}\}$.

6.15. Let A, B be rings, and let $A \otimes B$ be the tensor product of their additive groups.

(a) Show that $A \otimes B$ has a unique multiplication operation that turns it into a ring and has the property that $(a \otimes b)(a' \otimes b') = (aa') \otimes (bb')$ for all a, $a' \in A$, b, $b' \in B$.

(b) Suppose that A and B are commutative. Prove that for every commutative ring C there is a bijection $\mathrm{Rhom}(A \otimes B, C) \to \mathrm{Rhom}(A, C) \times \mathrm{Rhom}(B, C)$ that is functorial in C; here Rhom indicates sets of ring homomorphisms.

6.16. Let A, B be abelian groups, and let $A \circ B$ be their sum, as in Section 5.

(a) Prove that there is a natural group homomorphism $A \circ B \to A \times B$ with kernel $[A \circ B, A \circ B]$.

(b) Let $K = [A \circ B, A \circ B]$, and let $[K, A \circ B]$ be the subgroup of K generated by all commutators $[x, y]$ with $x \in K$ and $y \in A \circ B$. Prove that the map $A \times B \to K/[K, A \circ B]$ that sends (a, b) to the coset of the commutator $[a, b]$ is bilinear, and that it induces a group isomorphism $A \otimes B \to K/[K, A \circ B]$.

7. Categories

A *category* $\mathcal{C}$ consists of

(i) a class $\mathrm{Ob}\,\mathcal{C}$ (whose elements are commonly called the *objects* of $\mathcal{C}$);

(ii) for any two objects A, B of $\mathcal{C}$ a set $\mathrm{Mor}(A, B)$ (whose elements are called the *morphisms* from A to B in $\mathcal{C}$);

(iii) for any three objects A, B, and C of $\mathcal{C}$ a map

$$\mathrm{Mor}(B, C) \times \mathrm{Mor}(A, B) \to \mathrm{Mor}(A, C), \qquad (f, g) \mapsto f \circ g$$

(called *composition*);

(iv) for each object A of $\mathcal{C}$ an element $1_A \in \mathrm{Mor}(A, A)$ (the *identity* morphism of A), such that the following two conditions are satisfied: first, one has

$$(f \circ g) \circ h = f \circ (g \circ h)$$

whenever $f \in \mathrm{Mor}(C, D)$, $g \in \mathrm{Mor}(B, C)$, and $h \in \mathrm{Mor}(A, B)$ for objects A, B, C, D of $\mathcal{C}$; and secondly,

$$f \circ 1_A = f, \qquad 1_B \circ f = f$$

for each morphism $f \in \mathrm{Mor}(A, B)$ in $\mathcal{C}$.

One thinks of a morphism $f \in \mathrm{Mor}(A, B)$ as an 'arrow' from A to B, notation $A \xrightarrow{f} B$. In many, but not all, important examples of categories, the objects are sets with some sort of extra structure, and the arrows are maps between those sets that 'respect' that extra structure in a suitable sense. Category theory is sometimes condescendingly referred to as 'arrow science', and it is indeed worth remarking that the morphisms are more important than the objects. In fact, it is not difficult to give an equivalent set of axioms for categories in which only the morphisms occur, the objects then being encoded in the corresponding identity morphisms.

Studying category theory for its own sake is not considered fashionable. Nevertheless, 'categorical thinking'—in which the objects of study play a less prominent role than the 'maps' connecting them—turns out to be extremely profitable in many areas of mathematics, a notable example being algebraic geometry. It is surprising how much one can do purely by arguing about formal properties of arrows, and how one can simplify complicated arguments by phrasing them in this way. In addition, the fundamental notions from category theory form often a guide towards the formulation of the 'right' questions to ask in some mathematical theory. As an example, the reader may look at Section 5, where several important group-theoretical constructions were characterized by means of 'universal properties' that were formulated strictly in terms of arrows. In many mathematical theories the construction of objects with similar universal properties is a problem of central importance.

We shall usually write fg instead of $f \circ g$, and other notation is also found. Notational variants for $\mathrm{Mor}(A, B)$ include $\mathrm{Hom}(A, B)$, (A, B), $\mathrm{Mor}_{\mathcal{C}}(A, B)$, ${}_{\mathcal{C}}\mathrm{Hom}(A, B)$, and $\mathcal{C}(A, B)$.

Examples. (i) The category **Sets** of sets. The class of objects of **Sets** is the class of all sets, and $\mathrm{Mor}(A, B)$ consists of all maps $A \to B$. Composition is composition of maps, and 1_A is the identity map $A \to A$.

(ii) The category **Gr** of groups. Here $\mathrm{Ob}\,\mathbf{Gr}$ is the class of groups, $\mathrm{Mor}(A, B)$ is the set of group homomorphisms $A \to B$, and 1_A and $\circ$ are as in the previous example. The category **Ab** of abelian groups is a subcategory of **Gr**. Here we call a category $\mathcal{D}$ a *subcategory* of a category $\mathcal{C}$ if one has $\mathrm{Ob}\,\mathcal{D} \subset \mathrm{Ob}\,\mathcal{C}$ and $\mathrm{Mor}_{\mathcal{D}}(A, B) \subset \mathrm{Mor}_{\mathcal{C}}(A, B)$ for all A, $B \in \mathrm{Ob}\,\mathcal{D}$, composition in $\mathcal{D}$ is the same as in $\mathcal{C}$, and for all $A \in \mathrm{Ob}\,\mathcal{D}$ the identity morphism 1_A from $\mathcal{C}$ belongs to $\mathrm{Mor}_{\mathcal{D}}(A, A)$. If one has $\mathrm{Mor}_{\mathcal{D}}(A, B) = \mathrm{Mor}_{\mathcal{C}}(A, B)$ for all A, $B \in \mathrm{Ob}\,\mathcal{D}$, then one speaks of a *full* subcategory. Again, $\mathbf{Ab} \subset \mathbf{Gr}$ is an example.

(iii) The categories **Crg** and **Rg** of commutative rings and rings, respectively. Often, when one defines a category, one just specifies what the objects are, the definition of the morphisms then being more or less obvious. Thus, in the present case the morphisms are the ring homomorphisms. Likewise, 1_A and $\circ$ are usually as in the previous examples.

(iv) The category **Top** of topological spaces. Here the morphisms are the *continuous* maps. Since the composition of continuous maps is continuous, this is indeed a category.

(v) Let G be a group. An object of the category ${}_G\mathbf{Sets}$ of G-sets is a G-set, i. e., a set X equipped with an action of G on X. The morphisms are the maps respecting the G-action. Likewise, we have the category ${}_G\mathbf{sets}$ of finite G-sets, which is a full subcategory of ${}_G\mathbf{Sets}$.

(vi) For each field k, one has the category ${}_k\mathbf{Vs}$ of vector spaces over k, with the k-linear maps as morphisms. This notation, and the corresponding notation ${}_k\mathbf{vs}$ for the full subcategory of *finite dimensional* k-vector spaces, is not standard. There is in fact, in the literature, only little uniformity in the notation chosen for categories, and much creativity is often spent on inventing suggestive abbreviations.

(vii) For a ring R, one has the categories ${}_R\mathbf{Ab}$ and $\mathbf{Ab}_R$ of left and right R-modules, respectively, with the R-linear maps as morphisms (see Section 4, Exercises). If R is a field, one has ${}_R\mathbf{Ab} = {}_R\mathbf{Vs}$.

(viii) To give an example of a different nature, let X be a set with a partial ordering $\leq$. Then we get a category $\mathcal{C}$ by defining $X = \operatorname{Ob}\mathcal{C}$, letting $\operatorname{Mor}(x, y)$ be a set consisting of one element if $x \leq y$, and putting $\operatorname{Mor}(x, y) = \emptyset$ if $x \not\leq y$. Since $\leq$ is reflexive and transitive, this does give rise to a category, with uniquely defined compositions and identity morphisms.

(ix) Let k be a field. Define a category $\mathcal{C}$ by putting $\operatorname{Ob}\mathcal{C} = \{0, 1, 2, \ldots\} = \mathbf{Z}_{\geq 0}$ and taking $\operatorname{Mor}(m, n)$, for m, $n \in \mathbf{Z}_{\geq 0}$, equal to the set $M(n \times m, k)$ of $n \times m$ matrices with entries from k. Composition is matrix multiplication, and 1_n is the $n \times n$ identity matrix.

This category is resembles the category ${}_k\mathbf{vs}$ from Example (vi). Namely, we can define a map $F\colon \operatorname{Ob}\mathcal{C} \to \operatorname{Ob}{}_k\mathbf{vs}$ by putting $F(n) = k^n$, and, for m, $n \in \operatorname{Ob}\mathcal{C}$, a map $F_{mn}\colon \operatorname{Mor}_{\mathcal{C}}(m, n) \to {}_k\operatorname{Hom}(k^m, k^n)$ by letting $F_{mn}(A)$ be the map $x \mapsto Ax$. This is an example of a 'functor', which will be defined to mean that identity morphisms and composition are respected. It provides an 'equivalence' between $\mathcal{C}$ and ${}_k\mathbf{vs}$; the precise definition will be given later in this section, but in the present case it means that each finite dimensional k-vector space is isomorphic to one of the form k^n, and that each k-linear map $k^m \to k^n$ comes from a unique $n \times m$-matrix over k. In general, two categories will for practical purposes be considered 'the same' if there is an equivalence between them.

(x) The category **Grhom** of group homomorphisms is an example of a category in which the objects are *themselves* maps. An object of **Grhom** is a group homomorphism $f\colon G_1 \to G_2$ (or, if one wishes to be more formal, a triple (G_1, G_2, f) consisting of two groups G_1 and G_2 and a group homomorphism f between them). A *morphism* from an object $f\colon G_1 \to G_2$ to an object $g\colon H_1 \to H_2$ is a pair of group homomorphisms $(h_1\colon G_1 \to H_1,\ h_2\colon G_2 \to H_2)$ for which the diagram

$$\begin{tikzcd} G_1 \arrow[r, "f"] \arrow[d, "h_1"'] & G_2 \arrow[d, "h_2"] \\ H_1 \arrow[r, "g"] & H_2 \end{tikzcd}$$

commutes. Composition is defined by $(h_1, h_2)\circ(i_1, i_2) = (h_1 i_1, h_2 i_2)$. The reader may verify that in this way one does obtain a category. Groups play no special role in this example: for any category $\mathcal{C}$, one can similarly define the category of morphisms in $\mathcal{C}$.

The opposite category. If $\mathcal{C}$ is a category, then one defines the *opposite category* $\mathcal{C}^{\mathrm{opp}}$ by taking

$$\operatorname{Ob}\mathcal{C}^{\mathrm{opp}} = \operatorname{Ob}\mathcal{C}, \quad \operatorname{Mor}_{\mathcal{C}^{\mathrm{opp}}}(A, B) = \operatorname{Mor}_{\mathcal{C}}(B, A), \quad f \circ^{\mathrm{opp}} g = g \circ f$$

for all A, $B \in \operatorname{Ob}\mathcal{C}$ and all pairs f, g that need to be composed. Thus, one simply reverses the direction of all arrows. It is easily checked that $\mathcal{C}^{\mathrm{opp}}$ is indeed a category, with the same identity morphisms as $\mathcal{C}$, and that $(\mathcal{C}^{\mathrm{opp}})^{\mathrm{opp}} = \mathcal{C}$.

Isomorphisms. Let $\mathcal{C}$ be a category. A morphism $f\colon A \to B$ in $\mathcal{C}$ is called an *isomorphism* if there exists a morphism $g\colon B \to A$ such that $f \circ g = 1_B$ and $g \circ f = 1_A$. If g exists, it is called the *inverse* of f; note that it is unique, since if h is also an inverse of f then one has $h = h1_B = h(fg) = (hf)g = 1_A g = g$. Two objects A and B are called *isomorphic* (in $\mathcal{C}$) if there exists an isomorphism $A \to B$. An isomorphism from an object A to itself is called an *automorphism* of A; the set $\operatorname{Aut} A$ or $\operatorname{Aut}_{\mathcal{C}} A$ of automorphisms of A forms a group under composition.

The notions just defined are equivalent to the familiar ones for the category **Gr**. In **Sets**, the isomorphisms are the bijections, and the automorphisms are the permutations. We refer to Exercises 7.4 and 7.5 for the categorical analogues of injections and surjections, which are called *monomorphisms* and *epimorphisms*, respectively.

Point objects and copoint objects. Let $\mathcal{C}$ be a category. An object P of $\mathcal{C}$ is called a *point object* (or *terminal object*, or *universally attracting object*) of $\mathcal{C}$ if for all for all $A \in \operatorname{Ob}\mathcal{C}$ one has $\#\operatorname{Mor}(A, P) = 1$. An object Q is a *copoint object* (or *initial object*, or *universally repelling object*) of $\mathcal{C}$ if it is a point object in the opposite category; or, equivalently, if for all $A \in \operatorname{Ob}\mathcal{C}$ one has $\#\operatorname{Mor}(Q, A) = 1$.

For example, a group of order 1 is both a point object and a copoint object in **Gr**. In **Sets**, the point objects are the sets of cardinality 1, and the only copoint object is the empty set. In **Rg**, the zero ring is a point object and **Z** is a copoint object.

If P and P' are two point objects in a category, then there are unique arrows $f\colon P \to P'$ and $g\colon P' \to P$ in $\mathcal{C}$; and since there are also unique arrows $P' \to P'$ and $P \to P$, the composed arrows fg and gf are necessarily equal to $1_{P'}$ and 1_P, respectively, so that f and g are isomorphisms. This shows that a point object in a category, if one exists, is uniquely determined up to a unique isomorphism. The same applies to copoint objects.

Products. Let $\mathcal{C}$ be a category, and let A, $B \in \operatorname{Ob}\mathcal{C}$. An object C of $\mathcal{C}$ is called a *product* of A and B (in $\mathcal{C}$) if one is given a functorial system of bijections $\vartheta_X: \operatorname{Mor}(X, C) \to \operatorname{Mor}(X, A) \times \operatorname{Mor}(X, B)$, with X ranging over $\operatorname{Ob}\mathcal{C}$; the *functoriality* of this system means that for all X, $Y \in \operatorname{Ob}\mathcal{C}$ and all $f \in \operatorname{Mor}(X, Y)$ the diagram

$$\begin{tikzcd} \operatorname{Mor}(Y, C) \arrow[r, "\vartheta_Y"] \arrow[d, "f^*"'] & \operatorname{Mor}(Y, A) \times \operatorname{Mor}(Y, B) \arrow[d, "(f^*, f^*)"] \\ \operatorname{Mor}(X, C) \arrow[r, "\vartheta_X"] & \operatorname{Mor}(X, A) \times \operatorname{Mor}(X, B) \end{tikzcd}$$

commutes; here the vertical arrow f^* on the left maps g to gf, and the vertical arrow (f^*, f^*) on the right maps a pair (g_1, g_2) to $(g_1 f, g_2 f)$.

This definition may at first sight look a bit forbidding, but there is fortunately a more intelligible way of phrasing it. In order to do this, let A, B, C, and $(\vartheta_X)_{X \in \operatorname{Ob}\mathcal{C}}$ have the properties just stated, and let $\vartheta_C(1_C) = (\pi_1, \pi_2)$, where $\pi_1: C \to A$ and $\pi_2: C \to B$ are morphisms (the so-called *projection morphisms* or *projections*). Taking $Y = C$ one sees from the functoriality diagram that for all X the map $\vartheta_X: \operatorname{Mor}(X, C) \to \operatorname{Mor}(X, A) \times \operatorname{Mor}(X, B)$ sends f to $(\pi_1 f, \pi_2 f)$. Conversely, if one fixes any two morphisms $\pi_1: C \to A$ and $\pi_2: C \to B$, then the system of maps $\operatorname{Mor}(X, C) \to \operatorname{Mor}(X, A) \times \operatorname{Mor}(X, B)$ sending f to $(\pi_1 f, \pi_2 f)$ is indeed functorial. In this way one deduces that one may equivalently define a product of A and B as an object C of $\mathcal{C}$ that is equipped with morphisms $\pi_1: C \to A$ and $\pi_2: C \to B$, such that C, π_1, and π_2 have the following *universal property*: if X is any object of $\mathcal{C}$, and $(h_1: X \to A,\ h_2: X \to B)$ is an arbitrary pair of morphisms, then there exists a unique morphism $h: X \to C$ such that $h_1 = \pi_1 h$ and $h_2 = \pi_2 h$:

$$\begin{tikzcd} & X \arrow[dl, "h_1"'] \arrow[d, "h"] \arrow[dr, "h_2"] & \\ A & C \arrow[l, "\pi_1"] \arrow[r, "\pi_2"'] & B. \end{tikzcd}$$

Proposition 1. *The product is uniquely determined by the universal property up to a canonical isomorphism; that is, if C and C' are both products of A and B, with corresponding projection morphisms π_i and π_i', for $i = 1$, 2, then there exists a unique isomorphism $\varphi: C' \to C$ such that $\pi_1' = \pi_1 \circ \varphi$ and $\pi_2' = \pi_2 \circ \varphi$.*

Proof. Fixing $\mathcal{C}$, A, and B, we define the category $\mathcal{D}$ as follows. Objects of $\mathcal{D}$ are triples (X, h_1, h_2), where X is in $\operatorname{Ob}\mathcal{C}$ and $h_1: X \to A$ and $h_2: X \to B$ are morphisms in $\mathcal{C}$. A morphism from (X, h_1, h_2) to another such triple (Y, j_1, j_2) is by definition a morphism $g: X \to Y$ in $\mathcal{C}$ such that $h_1 = j_1 g$ and $h_2 = j_2 g$. One readily verifies that in this manner $\mathcal{D}$ becomes a category. One also readily verifies that a triple (C, π_1, π_2) gives rise to a product of A and B in $\mathcal{C}$ if and only if that triple is a point object in $\mathcal{D}$. Above we proved the uniqueness of point objects, up to a unique isomorphism, and this implies the proposition.

The reader is encouraged to see what the proof just given looks like if it is phrased entirely in terms of the category $\mathcal{C}$, without invoking $\mathcal{D}$.

A different proof of Proposition 1, along the lines of the arguments given in Section 5, depends on *Yoneda's lemma* (see Exercise 7.33(b)), which is the analogue, for general categories, of the group-theoretical Theorem 3 of Section 5 (see also Exercise 5.6). Yoneda's lemma implies, roughly speaking, that an object C of a category is uniquely determined, up to isomorphism, by the 'family of sets of incoming arrows' $\mathrm{Mor}(X, C)$, with X ranging over $\mathrm{Ob}\,\mathcal{C}$; the precise statement again involves a functoriality condition. If C and C' are both products of A and B, then $\mathrm{Mor}(X, C)$ and $\mathrm{Mor}(X, C')$ can both be identified with $\mathrm{Mor}(X, A) \times \mathrm{Mor}(X, B)$, and hence with each other, so that by Yoneda's lemma C is isomorphic to C'. The details of this argument are left to the reader.

Since a product of A and B, if it exists, is uniquely determined up to isomorphism, one calls it *the* product of A and B. It is denoted by $A \times B$; one also finds the notation $A \amalg B$. In many categories, such as **Sets**, **Gr**, **Rg**, and **Top**, one can indeed take $A \times B$ to be the ordinary cartesian product of A and B (with componentwise operations, and in the case of **Top** with the product topology), and the π_i the usual projection maps.

One can also define products of arbitrary collections of objects. If $(A_i)_{i \in I}$ is a collection of objects in a category $\mathcal{C}$, then a *product* of the A_i is an object C of $\mathcal{C}$ together with a collection of morphisms $\pi_i: C \to A_i$, for $i \in I$, such that for any object X of $\mathcal{C}$ and any family $(h_i: X \to A_i)_{i \in I}$ of morphisms in $\mathcal{C}$ there exists a unique morphism $h: X \to C$ such that for all $i \in I$ one has $h_i = \pi_i h$. By the same argument as for products of two objects, one proves that a product is unique up to a unique isomorphism, if it exists. In **Sets**, **Gr**, **Rg**, and **Top** the products are again the usual ones. An 'empty' product—that is, a product with $I = \emptyset$—is the same as a point object.

Sums. The notion dual to 'product' is called *sum*; that is, something is a sum in $\mathcal{C}$ if it is a product in the opposite category $\mathcal{C}^{\mathrm{opp}}$. Thus, if A and B are objects in a category $\mathcal{C}$, then a *sum* of A and B in $\mathcal{C}$ is an object C of $\mathcal{C}$ equipped with two morphisms $i_1: A \to C$ and $i_2: B \to C$, such that for any object X and any pair of morphisms $h_1: A \to X$, $h_2: B \to X$, there is a unique morphism $h: C \to X$ such that $h_1 = h i_1$ and $h_2 = h i_2$:

$$\begin{tikzcd} & X & \\ A \arrow[ur, "h_1"] \arrow[r, "i_1"'] & C \arrow[u, "h"] & B. \arrow[ul, "h_2"'] \arrow[l, "i_2"] \end{tikzcd}$$

Similarly, one can define sums of arbitrary collections of objects. Sums are unique in the same sense as products are.

For the sum of A and B one finds the notation $A \Sigma B$; instead of 'sum' one also uses the term *coproduct*, with corresponding notation $A \amalg B$.

In **Sets**, sums are disjoint unions. In **Gr**, the sum of A and B is the group $A \circ B$ constructed in Section 5. In the category **Crg** of commutative rings, the sum is the tensor product (see Exercise 6.15). In **Ab**, the sum of A and B is the same as the product $A \times B$, the maps $i_1 \colon A \to A \times B$ and $i_2 \colon B \to A \times B$ being defined by $i_1(a) = (a, 0)$ and $i_2(b) = (0, b)$; one also calls $A \times B$ the *direct sum* of A and B, notation $A \oplus B$. The sum of an arbitrary collection $(A_i)_{i \in I}$ of objects in **Ab** is given by

$$\bigoplus_{i \in I} A_i = \big\{(a_i)_{i \in I} \in \prod_{i \in I} A_i : \#\{i \in I : a_i \neq 0\} < \infty\big\}.$$

Note that this is different from the full product $\prod_{i \in I} A_i$ if there are infinitely many $i \in I$ for which A_i has order greater than 1.

In the exercises one finds variants of the notion of a products: *fibred products* in Exercise 7.11 and *left limits* in Exercise 7.14. The dual notions are *fibred sums* and *right limits.*

Functors. We next extend our 'categorical thinking' to categories themselves, and define the 'morphisms' *between* categories. They are called *functors.*

Let $\mathcal{C}$ and $\mathcal{D}$ be categories. A *functor* or *covariant functor* $F \colon \mathcal{C} \to \mathcal{D}$ consists of a map $F \colon \mathrm{Ob}\,\mathcal{C} \to \mathrm{Ob}\,\mathcal{D}$ as well as, for any two objects A and B of $\mathcal{C}$, a map $F_{AB} \colon \mathrm{Mor}_{\mathcal{C}}(A, B) \to \mathrm{Mor}_{\mathcal{D}}(F(A), F(B))$, with the property that composition and identities are preserved:

$$F_{AC}(f \circ g) = F_{BC}(f) \circ F_{AB}(g), \qquad F_{AA}(1_A) = 1_{F(A)}$$

(whenever meaningful). A *contravariant functor* $F \colon \mathcal{C} \to \mathcal{D}$ is a covariant functor $\mathcal{C} \to \mathcal{D}^{\mathrm{opp}}$ (or, equivalently, a functor $\mathcal{C}^{\mathrm{opp}} \to \mathcal{D}$).

Just as with morphisms, one can compose functors to get other functors, functor composition is associative, and there are obvious identity functors. Note that the composition of two contravariant functors is covariant.

It is tempting to conclude that one can now talk about the category **Cat** of all categories, the objects being categories and the morphisms being covariant functors; however, in this manner one gets dangerously close to the paradoxes of set theory, which involve objects (such as the collection of all sets x with $x \notin x$) that are too 'large' to be called 'sets'. If one does want to talk about a category of categories one needs to take special precautions. One possible precaution would be to include in **Cat** only categories that are *small* in the sense that their classes of objects are *sets.* If one finds this too restrictive—since it excludes many interesting categories such as **Sets** and **Gr**—then another option is to find a way of adapting the axiomatic set theory one is working in to the needs of category theory. In the practice of working with categories it is hardly ever necessary to bother about these issues, but it is good to be aware of their existence.

Examples. (xi) The functor $\mathbf{Gr} \to \mathbf{Sets}$ sending a group G to its underlying set and each homomorphism to itself (viewed as a map) is an example of a functor. Such a functor, in which part of the structure is 'forgotten', is called a *forgetful functor.*

(xii) The map $\mathbf{Rg} \to \mathbf{Gr}$ that sends each ring R to its unit group R^* extends naturally to a functor, since each ring homomorphism $R \to R'$ maps R^* to R'^* and therefore induces a group homomorphism $R^* \to R'^*$. In many cases, one indicates a functor just by saying what it does on objects, its effect on morphisms then being more or less obvious.

(xiii) The map $F\colon \mathbf{Gr} \to \mathbf{Rg}$ sending G to the group ring $\mathbf{Z}[G]$ and a group homomorphism $f\colon G \to H$ to the ring homomorphism $F_{GH}(f)\colon \mathbf{Z}[G] \to \mathbf{Z}[H]$ defined by $\sum_g n_g g = \sum_g n_g f(g)$ is a functor.

(xiv) The inclusion $\mathbf{Ab} \subset \mathbf{Gr}$ may be viewed as a functor, the *inclusion* functor. In the opposite direction, one has the *abelianization* functor $\mathbf{Gr} \to \mathbf{Ab}$ sending G to $G^{\mathrm{ab}} = G/[G, G]$.

(xv) A functor $\mathcal{C} \to \mathcal{D}$ is called *constant* if it maps all objects of $\mathcal{C}$ to the same object D of $\mathcal{D}$, and all morphisms to 1_D. A constant functor is both covariant and contravariant.

(xvi) Let k be a field. Then the map $D\colon {}_k\mathbf{Vs} \to {}_k\mathbf{Vs}$ sending V to $V^* = \mathrm{Hom}_k(V, k)$ is an example of a contravariant functor; it sends $f\colon V \to W$ to the map $f^*\colon W^* \to V^*$ sending λ to $\lambda \circ f$.

(xvii) As an example of a construction that is *not* a functor we mention the algebraic closure of fields. One can show that there does not exist a functor from the category of fields to itself that sends each field k to an algebraic closure $\bar{k}$ of k and each field homomorphism $f\colon k \to l$ to a field homomorphism $\bar{k} \to \bar{l}$ that extends f (see Exercise 7.23). Each f does have such an extension, but these extensions cannot be simultaneously constructed in such a way that the rule $F(f_1 f_2) = F(f_1)F(f_2)$ is satisfied.

(xviii) In algebraic topology, one constructs for each pathwise connected topological space X a group $\pi(X)$ that is called the *fundamental group* of X. Its definition involves a choice that makes it impossible to extend it to a functor from the category $\mathbf{Ptop}$ of such topological spaces to $\mathbf{Gr}$. To remedy this, one defines instead the category $\mathbf{Top}^*$ of topological spaces with 'base point'; an object of $\mathbf{Top}^*$ is a pair (X, x) consisting of a topological space X and a point $x \in X$, and a morphism $(X, x) \to (Y, y)$ is a continuous map $f\colon X \to Y$ satisfying $f(x) = y$. Now there is a functor $\pi\colon \mathbf{Top}^* \to \mathbf{Gr}$ sending (X, x) to the group $\pi(X, x)$ of homotopy classes of paths in X from x to x. Another way to remedy the problem is to pass from $\mathbf{Gr}$ to $\mathbf{Ab}$; one can show that there is indeed a functor $\mathbf{Ptop} \to \mathbf{Ab}$ sending each X to $\pi(X)^{\mathrm{ab}}$.

(xix) The map $F\colon \mathbf{Crg} \to \mathbf{Sets}$ sending R to $\{(x, y) \in R \times R : x^2 + y^2 = 1\}$ is a functor. If we provide $F(R)$ with an abelian group structure by putting $(x,y)+(x',y') = (xx'-yy', xy'+x'y)$ then we obtain a functor $\mathbf{Crg} \to \mathbf{Ab}$ that composed with the forgetful functor $\mathbf{Ab} \to \mathbf{Sets}$ yields F.

(xx) Let $\mathcal{C}$ be a category, and fix an object X of $\mathcal{C}$. Then there is a covariant functor $h_X: \mathcal{C} \to \mathbf{Sets}$ sending Y to $\mathrm{Mor}(X,Y)$ and sending $f \in \mathrm{Mor}(Y,Z)$ to the map $f_*: \mathrm{Mor}(X,Y) \to \mathrm{Mor}(X,Z)$ that maps g to fg. Likewise, there is a contravariant functor $h^X: \mathcal{C} \to \mathbf{Sets}$ sending Y to $\mathrm{Mor}(Y,X)$ and f to f^*, where f^* is 'composition with f on the right'. One calls the functors h_X and h^X the functors *represented* by X.

(xxi) Consider the functor $h_{\mathbf{Z}}: \mathbf{Gr} \to \mathbf{Sets}$ from the previous example. It sends a group G to $\mathrm{Hom}(\mathbf{Z}, G)$. The latter set can for all practical purposes be identified with the underlying set of G itself, via the bijection $\mathrm{Hom}(\mathbf{Z}, G) \to G$ sending f to $f(1)$. Thus, $h_{\mathbf{Z}}$ is in a sense 'the same' as the forgetful functor $\mathbf{Gr} \to \mathbf{Sets}$. To formulate this precisely, we shall turn the collection of all functors $\mathbf{Gr} \to \mathbf{Sets}$ into a category; in that category, $h_{\mathbf{Z}}$ and the forgetful functor are *isomorphic*.

Functoriality and functor categories. Let $\mathcal{C}$ and $\mathcal{D}$ be categories, and let E, $F: \mathcal{C} \to \mathcal{D}$ be covariant functors. Suppose that for each object A of $\mathcal{C}$ one is given a morphism $\vartheta_A: E(A) \to F(A)$ in $\mathcal{D}$. The system $(\vartheta_A)_{A \in \mathrm{Ob}\,\mathcal{C}}$ is said to be *functorial in A* if for every morphism $f: A \to B$ in $\mathcal{C}$ the diagram

$$\begin{array}{ccc} E(A) & \xrightarrow{\vartheta_A} & F(A) \\ {\scriptstyle E(f)}\big\downarrow & & \big\downarrow{\scriptstyle F(f)} \\ E(B) & \xrightarrow{\vartheta_B} & F(B) \end{array}$$

commutes in $\mathcal{D}$. In special cases, we saw this notion already in Section 5, and in the definition of products given above. By a *morphism $\vartheta: E \to F$ of functors*, also called a *natural transformation* from E to F, we mean a system $\vartheta = (\vartheta_A)_{A \in \mathrm{Ob}\,\mathcal{C}}$ of morphisms $\vartheta_A: E(A) \to F(A)$ in $\mathcal{D}$ that is functorial in A. For example, if $E = F$ then $\vartheta_A = 1_{E(A)}$ yields the *identity* morphism of E. Morphisms $D \to E$ and $E \to F$ can be composed in an obvious way to yield a morphism $D \to F$, and composition is associative. One might conclude that the collection of all covariant functors $\mathcal{C} \to \mathcal{D}$ is itself the collection of objects of a category, the *category of functors from $\mathcal{C}$ to $\mathcal{D}$*, the morphisms being as just defined; but here again one runs into set-theoretical difficulties, since the collection of morphisms between two given functors may be too large to be called a 'set'. We shall be a little cavalier about this, and pretend that this problem does not exist, or that it has been successfully addressed by people working in the foundations of mathematics.

Examples. (xxii) Let $F: \mathbf{Crg} \to \mathbf{Sets}$ be the functor from Example (xix), defined by $F(R) = \{(x,y) \in R \times R : x^2 + y^2 = 1\}$. The map $\vartheta_R: F(R) \to R$ defined by $\vartheta(x,y) = x$ is functorial in R, so it gives rise to a morphism from F to the forgetful functor $\mathbf{Crg} \to \mathbf{Sets}$. If, as in

Example (xix), we turn F into a functor $\mathbf{Crg} \to \mathbf{Ab}$, then there is a morphism $\eta: F \to F$ defined by $\eta_R(x,y) = (x,-y)$. Since $\eta^2 = 1_{F(R)}$, this is an example of an *automorphism* of F.

(xxiii) The functorial bijection $\mathrm{Hom}(\mathbf{Z}, G) \to G$, $f \mapsto f(1)$, from Example (xxi) above defines a morphism from $h_{\mathbf{Z}}$ to the forgetful functor. It is actually an *isomorphism* of functors, that is, it is an isomorphism in the category of functors $\mathbf{Gr} \to \mathbf{Sets}$. (In general, a morphism $\vartheta = (\vartheta_A)_{A \in \mathrm{Ob}\,\mathcal{C}}$ of functors $\mathcal{C} \to \mathcal{D}$ is an isomorphism if and only if each ϑ_A is an isomorphism in $\mathcal{D}$.) Thus, the forgetful functor $\mathbf{Gr} \to \mathbf{Sets}$ is isomorphic to $h_{\mathbf{Z}}$. In general, one calls a covariant (or contravariant) functor $F: \mathcal{C} \to \mathbf{Sets}$ *representable* if there is an object X of $\mathcal{C}$ such that F is isomorphic to h_X (or h^X). So the forgetful functor $\mathbf{Gr} \to \mathbf{Sets}$ is representable. Also the functor $\mathbf{Crg} \to \mathbf{Sets}$ from Example (xix) is representable, a representing object being $\mathbf{Z}[X,Y]/(X^2+Y^2-1)$.

(xxiv) Let k be a field and let $D: {}_k\mathbf{Vs} \to {}_k\mathbf{Vs}$ be the contravariant functor from Example (xvi); it sends V to $V^* = \mathrm{Hom}_k(V,k)$. Then $D^2 = D \circ D$ is a covariant functor ${}_k\mathbf{Vs} \to {}_k\mathbf{Vs}$. Denote by $\mathrm{id}: {}_k\mathbf{Vs} \to {}_k\mathbf{Vs}$ the identity functor. A morphism $\vartheta: \mathrm{id} \to D^2$ is given by $\vartheta_V(v)(\lambda) = \lambda(v)$, for $v \in V$ and $\lambda \in V^*$. If both functors are restricted to the full subcategory ${}_k\mathbf{vs}$ of ${}_k\mathbf{Vs}$, then ϑ becomes an isomorphism of functors.

(xxv) Let X and Y be objects of a category $\mathcal{C}$, and let $f \in \mathrm{Mor}(X,Y)$. Then there is a morphism $f^*: h_Y \to h_X$ that sends any $g \in h_Y(A) = \mathrm{Mor}(Y,A)$ to $f^*(g) = g \circ f \in h_X(A) = \mathrm{Mor}_{\mathcal{C}}(X,A)$, for any $A \in \mathrm{Ob}\,\mathcal{C}$. *Yoneda's lemma* (see Exercise 7.33(b)) asserts that the map $\mathrm{Mor}_{\mathcal{C}}(X,Y) \to \mathrm{Mor}(h_Y, h_X)$ sending f to f^* is a *bijection.*

Equivalences of categories. Let $\mathcal{C}$ and $\mathcal{D}$ be categories. A covariant functor $F: \mathcal{C} \to \mathcal{D}$ is called an *equivalence*, or an *equivalence of categories*, if there is a covariant functor $G: \mathcal{D} \to \mathcal{C}$ such that FG is isomorphic to the identity functor of $\mathcal{D}$ and GF is isomorphic to the identity functor of $\mathcal{C}$. A contravariant functor $F: \mathcal{C} \to \mathcal{D}$ is called an *anti-equivalence* if it is an equivalence $\mathcal{C} \to \mathcal{D}^{\mathrm{opp}}$. The categories $\mathcal{C}$ and $\mathcal{D}$ are called *equivalent* (or *anti-equivalent*) if there is an equivalence (or anti-equivalence) $\mathcal{C} \to \mathcal{D}$.

Note that we are not requiring that FG and GF are *equal* to the identity functors of $\mathcal{D}$ and $\mathcal{C}$; that requirement would define the notion of an *isomorphism* of categories. For most applications of categories, the latter notion is not nearly as important as the notion of equivalence of categories. This is because in most categories one really cares about objects only 'up to isomorphism'; accordingly, one does not require that FG and GF send each object to *itself*, but one is, roughly speaking, satisfied if they send each object to an *isomorphic* object.

An important sufficient condition for a functor to be an equivalence is found in Exercise 7.28.

Examples. (xxvi) The contravariant functor $D: {}_k\mathbf{vs} \to {}_k\mathbf{vs}$ from Example (xxiv) is an anti-equivalence of ${}_k\mathbf{vs}$ with itself, since D^2 is isomorphic to the identity functor ${}_k\mathbf{vs} \to {}_k\mathbf{vs}$.

(xxvii) Let k be a field, and let the category $\mathcal{C}$ and the functor $F: \mathcal{C} \to {}_k\mathbf{vs}$ be as defined in Example (ix). The criterion from Exercise 7.28 shows that F is an equivalence.

(xxviii) Let $\mathcal{C}$ be a category, and let $\mathbf{Sets}^{\mathcal{C}}$ be the category of covariant functors $\mathcal{C} \to \mathbf{Sets}$. From Yoneda's lemma (see Example (xxv)) one deduces that the contravariant functor $\mathcal{C} \to \mathbf{Sets}^{\mathcal{C}}$ sending an object X to h_X and a morphism f to f^* yields an anti-equivalence between $\mathcal{C}$ and the full subcategory of $\mathbf{Sets}^{\mathcal{C}}$ consisting of the representable functors. Likewise, the functor sending X to h^X is an equivalence of $\mathcal{C}$ with the category of contravariant representable functors $\mathcal{C} \to \mathbf{Sets}$, the morphisms in the latter category being morphisms of functors.

7. Exercises

7.1. Let $\mathcal{C}$ be a category, and let $A \in \operatorname{Ob} \mathcal{C}$. The category of *objects over* A, or of *morphisms to* A, has as its objects pairs (X, f) consisting of an object X of $\mathcal{C}$ and a morphism $f: X \to A$ in $\mathcal{C}$, a morphism $(X, f) \to (Y, g)$ being a morphism $h: X \to Y$ in $\mathcal{C}$ with $f = gh$. Prove that this is indeed a category, and that it has a point object.

7.2. (a) A *rng* is defined by omitting, in the definition of 'ring', the requirement that there be a unit element, and a *rng homomorphism* is a map preserving addition and multiplication. Prove that the category $\mathbf{Rg}$ of rings is a subcategory of the category $\mathbf{Rng}$ of rngs, but not a full one.

(b) Let $\mathcal{C}$ be a full subcategory of a category $\mathcal{D}$, and let A, $B \in \operatorname{Ob} \mathcal{C}$. Prove: A and B are isomorphic in $\mathcal{C}$ if and only if they are isomorphic in $\mathcal{D}$.

7.3. Suppose that

$$\begin{array}{ccc} \bullet & \rightarrow & \bullet \\ \downarrow & \swarrow & \downarrow \\ \bullet & \rightarrow & \bullet \end{array}$$

is a commutative diagram in a category, with the property that both vertical arrows are isomorphisms. Prove that the other three arrows are isomorphisms as well.

7.4. A morphism f in a category is called a *monomorphism* if it has the property that $fg = fh \Rightarrow g = h$ for all morphisms g, h for which fg and fh are defined. Prove that in $\mathbf{Sets}$, $\mathbf{Gr}$, $\mathbf{Rg}$, $\mathbf{Top}$, and in the category of fields each injective morphism is a monomorphism, and decide for each of these categories whether conversely every monomorphism is injective.

7.5. A morphism f in a category is called an *epimorphism* if it is a monomorphism in the opposite category. Prove that in $\mathbf{Sets}$, $\mathbf{Gr}$, $\mathbf{Rg}$, $\mathbf{Top}$, and in the category of fields each surjective morphism is an epimorphism, and decide for each of these categories whether conversely every epimorphism is surjective. (*Hint* in the case of groups: prove that if $H \neq G$ is a subgroup of a group G, then the disjoint union of two copies of G/H, viewed as a G-set, has an automorphism as an H-set that is not an automorphism as a G-set.)

7.6. Let $f: A \to B$ be a morphism in a category $\mathcal{C}$.

(a) Prove: f is a monomorphism if and only if for each object X in $\mathcal{C}$ the map $f_*: \mathrm{Mor}(X, A) \to \mathrm{Mor}(X, B)$ sending g to fg is injective.

(b) Prove: f is an epimorphism if for each object X in $\mathcal{C}$ the map $f_*: \mathrm{Mor}(X, A) \to \mathrm{Mor}(X, B)$ sending g to fg is surjective; prove also that the converse ("only if") is correct in the category of sets, but not in the category of groups.

7.7. Decide for each of the following categories whether it has a point object, and prove the correctness of your assertions: **Sets**, **Gr**, **Rg**, **Top**, and the category of fields.

7.8. Decide for each of the following categories whether it has a copoint object, and prove the correctness of your assertions: **Sets**, **Gr**, **Rg**, **Top**, the category of finite rings, the category of fields, and the category of fields of characteristic 7.

7.9. Prove that in **Crg** any collection of objects has a product and a sum.

7.10. Let $\mathbf{C}$ be the field of complex numbers, viewed as an object of the category $\mathcal{C}$ of fields. Prove that in $\mathcal{C}$ there is neither a product nor a sum of $\mathbf{C}$ with itself.

7.11. Let $\mathcal{C}$ be a category, A, B, and C objects of $\mathcal{C}$, and let $f: B \to A$ and $g: C \to A$ be morphisms. By a *fibred product* (or a *pull-back*) of B and C over A one means a pair (X, h) consisting of an object X of $\mathcal{C}$ and a morphism $h: X \to A$ such that (X, h) is a product of (B, f) and (C, g) in the category of objects over A (see Exercise 7.1). If a fibred product (X, h) exists, one often writes $X = B \times_A C$, the morphisms f, g, and h being understood.

(a) Formulate the definition just given directly in terms of $\mathcal{C}$.

(b) Prove that fibred products exist in **Sets**, **Gr**, and **Rg**.

7.12. Dually to fibred products, one has *fibred sums* (or *push-outs*). Give the definition of fibred sums yourself, and prove that they exist in **Sets**, **Gr**, and **Crg**. *Do they exist in **Rg**?

7.13. By a *graph* we mean a quadruple $\Gamma = (V, E, s, t)$ where V and E are sets and s and t are maps; one should think of the elements of V as *vertices*, and of each element e of E as a directed *edge* going from its *source* $s(e)$ to its *target* $t(e)$. Let $\Gamma = (V, E, s, t)$ be a graph, and let $\mathcal{C}$ be a category. A *diagram* of *shape* Γ in $\mathcal{C}$ is a collection $(A_v)_{v \in V}$ of objects in $\mathcal{C}$ (one object A_v for each $v \in V$), together with a collection $(f_e)_{e \in E}$ of morphisms in $\mathcal{C}$, with $f_e \in \mathrm{Mor}(A_{s(e)}, A_{t(e)})$ for each $e \in E$.

(a) Think, for fixed Γ and $\mathcal{C}$, of a natural notion of *morphisms* between diagrams of shape Γ in $\mathcal{C}$ in such a manner that one obtains a category. Your definition should be such that in the case $\#V = 1$, $E = \emptyset$ one recovers the category $\mathcal{C}$, and that in the case $\mathcal{C} = \mathbf{Gr}$, $\#V = 2$, $\#E = 1$, $V = (sE) \cup (tE)$ one recovers the category **Grhom** defined in Example (x).

(b) Define what it would mean for a diagram to be *commutative*. Which graph Γ should be chosen so that one recovers the notion of a commutative square?

7.14. Let $\Gamma = (V, E, s, t)$ and $\mathcal{C}$ be as in Exercise 7.13, and let $(A_v)_{v\in V}$ and $(f_e)_{e\in E}$ constitute a diagram of shape Γ in $\mathcal{C}$, as in Exercise 7.13. A *morphism* g from an object B of $\mathcal{C}$ to this diagram is a collection of morphisms $g_v \in \mathrm{Mor}(B, A_v)$, one for each $v \in V$, such that for all $e \in E$ one has $g_{t(e)} = f_e g_{s(e)}$. A *left limit* (or *limit*, or *inverse limit*) of the diagram $\big((A_v)_{v\in V}, (f_e)_{e\in E}\big)$ is an object L of $\mathcal{C}$, together with a morphism $l = (l_v)_{v\in V}$ from L to the diagram, such that for every object B of $\mathcal{C}$ and every morphism $g = (g_v)_{v\in V}$ from B to the diagram there exists a unique $h \in \mathrm{Mor}(B, L)$ with $g = lh$ (that is: $g_v = l_v h$ for all $v \in V$).

Construct a category whose objects are all pairs (B, g), where B is an object of $\mathcal{C}$ and g is a morphism from B to the given diagram. Do this in such a way that a point object in that category is the same as a left limit of the given diagram.

7.15. (a) Show that a left limit, if it exists, is unique up to a unique isomorphism.

(b) Show that the notions of *product* (of any number of objects), *fibred product* (see Exercise 7.11), and *point object* are special cases of left limits.

7.16. (a) Show that left limits exist in **Sets**, in **Gr**, in **Ab**, and in **Rg**. Do they exist in the category of *finite* sets? and in the category of fields?

(b) Show that kernels are special cases of left limits, both in **Gr** and in **Ab**.

7.17. Let $\Gamma = (V, E, s, t)$ be a graph (see Exercise 7.13). Interchanging s and t (i.e., reversing the direction of all edges), we obtain the graph $\Gamma^{\mathrm{opp}} = (V, E, t, s)$ *opposite* to Γ. Every diagram of shape Γ in $\mathcal{C}$ is a diagram of shape Γ^{opp} in $\mathcal{C}^{\mathrm{opp}}$. A left limit of that diagram in $\mathcal{C}^{\mathrm{opp}}$ is called a *right limit* (or *colimit*, or *direct limit*) of the original diagram in $\mathcal{C}$.

Do the analogues of Exercises 7.15 and 7.16 for right limits. Pay special attention to the existence of cokernels (the dual of kernels) in **Gr**.

7.18. Let $f\colon A \to B$ be a morphism in a category, and consider the diagram $A \xrightarrow{f} B \xleftarrow{f} A$. Prove: f is a monomorphism if and only if the object A, together with the morphisms $A \xrightarrow{1} A$, $A \xrightarrow{f} B$, $A \xrightarrow{1} A$ (where $1 = 1_A$), is a left limit of this diagram. What is the corresponding characterization of epimorphisms?

7.19. In Section 6, we defined for every set S a group $F(S)$. Extend F to a functor **Sets** $\to$ **Gr**.

7.20. Let S be a set.

(a) Construct a commutative ring $A(S)$ together with a (set-theoretic) map $i\colon S \to A(S)$ such that for every pair consisting of a commutative ring R and a map $f\colon S \to R$

there exists a unique ring homomorphism $g: A(S) \to R$ with $f = g \circ i$. Show also that the assignment $S \mapsto A(S)$ can be extended to a functor **Sets** $\to$ **Crg**.

(b) As (a), but with the word "commutative" (twice) left out, and **Crg** replaced by **Rg**.

7.21. Show that in **Rg** any collection of objects has a sum. *Hint*: use Exercise 7.20(b).

7.22. Let S be a set.

(a) Construct a topological space $X(S)$ together with a map $i: S \to X(S)$ such that for every pair consisting of a topological space Z and a map $f: S \to Z$ there exists a unique continuous map $g: X(S) \to Z$ with $f = g \circ i$.

(b) Construct a topological space $Y(S)$ together with a map $i: Y(S) \to S$ such that for every pair consisting of a topological space Z and a map $f: Z \to S$ there exists a unique continuous map $g: Z \to Y(S)$ with $f = i \circ g$.

(c) As in Exercises 7.19 and 7.20, extend X and Y to functors **Sets** $\to$ **Top**.

7.23. (a) Let $K \subset L \subset M$ be fields with $[L : K] = 2$ and $[M : L] = 2$, and suppose that M is a Galois extension of K with a cyclic Galois group. Let $\bar{L}$ be an algebraic closure of L. Prove that there is no field automorphism σ of $\bar{L}$ of order 2 such that the restriction of σ to L generates the Galois group of L over K.

(b) Prove that there does not exist a functor from the category of fields to itself that sends each field k to an algebraic closure $\bar{k}$ of k and each field homomorphism $f: k \to l$ to a field homomorphism $\bar{k} \to \bar{l}$ that extends f.

7.24. Show that for any three abelian groups A, B, C there is a group isomorphism $f_{ABC}: \mathrm{Hom}(A \otimes B, C) \to \mathrm{Hom}\big(A, \mathrm{Hom}(B, C)\big)$. Formulate what it would mean to say that the system of maps f_{ABC} is "functorial" in all three "variables" A, B, C simultaneously. Prove that the maps f_{ABC} that you constructed do have that property.

7.25. Define the functor U: **Rg** $\to$ **Sets** by letting $U(R)$ be the set of units of R. Let A be the ring $\mathbf{Z}[X, Y]/(XY - 1)$, and h_A: **Rg** $\to$ **Sets** the functor represented by A. Exhibit an isomorphism of functors $U \cong h_A$.

7.26. Let F: **Crg** $\to$ **Sets** be the functor defined in Example (xix). Verify that F is indeed a functor, and prove the correctness of all assertions made about this functor in Examples (xix), (xxii), and (xxiii).

7.27. Let $F: \mathcal{C} \to \mathcal{D}$ be an equivalence of categories $\mathcal{C}$ and $\mathcal{D}$. Prove: if A, B are objects of $\mathcal{C}$, then one has $A \cong_{\mathcal{C}} B$ if and only if $F(A) \cong_{\mathcal{D}} F(B)$.

7.28. Let $\mathcal{C}$ and $\mathcal{D}$ be categories, and let $F: \mathcal{C} \to \mathcal{D}$ be a covariant functor. Prove: F is an equivalence if and only if (i) for every object X of $\mathcal{D}$ there is an object A of $\mathcal{C}$

with $X \cong_{\mathcal{D}} F(A)$, and (ii) for every two objects A, B of $\mathcal{C}$, the map $F\colon \mathrm{Mor}_{\mathcal{C}}(A,B) \to \mathrm{Mor}_{\mathcal{D}}(F(A),F(B))$ is bijective.

7.29. Prove that every category $\mathcal{C}$ has a full subcategory $\mathcal{D}$ such that (i) the inclusion $\mathcal{D} \subset \mathcal{C}$ is an equivalence of categories; (ii) for any two objects A, B of $\mathcal{D}$ with $A \cong_{\mathcal{D}} B$ one has $A = B$.

7.30. Prove the assertion made in Example (xxvii).

7.31. Let k be a field, let $D\colon {}_k\mathbf{Vs} \to {}_k\mathbf{Vs}$ be the functor defined in Example (xvi), and let $\vartheta = (\vartheta_V)_V\colon \mathrm{id} \to D^2$ be the morphism of functors from Example (xxiv). Define the morphisms $\zeta\colon D \to D^3$ and $\eta\colon D^3 \to D$ of functors by $\zeta_V = \vartheta_{D(V)}$ and $\eta_V = D(\vartheta_V)$. Prove that the morphism $\eta \circ \zeta\colon D \to D$ is the identity morphism of D.

7.32. Let **Fab** be the category of finite abelian groups.

(a) Show that for each finite abelian group A the group $D(A) = \mathrm{Hom}(A, \mathbf{Q}/\mathbf{Z})$ is finite and abelian, and that D extends in a natural way to a contravariant functor $D\colon \mathbf{Fab} \to \mathbf{Fab}$.

(b) Prove that the covariant functors $1_{\mathbf{Fab}}$, $D^2\colon \mathbf{Fab} \to \mathbf{Fab}$ are isomorphic, and that D is an anti-equivalence of categories.

(c) Prove that for every finite abelian group A one has $A \cong D(A)$. Does it make sense to ask whether this isomorphism is functorial in A?

7.33. Let $\mathcal{C}$ be a category.

(a) Verify the statement, implicitly made in Example (xxviii), that there is a contravariant functor $h\colon \mathcal{C} \to \mathbf{Sets}^{\mathcal{C}}$ sending an object X to h_X and a morphism f to f^*.

(b) (*Yoneda's lemma.*) Let X, $Y \in \mathrm{Ob}\,\mathcal{C}$, and write $\mathrm{Mor}(h_Y, h_X)$ for the set of morphisms $h_Y \to h_X$ of functors. Yoneda's lemma asserts that *the map* $\mathrm{Mor}_{\mathcal{C}}(X,Y) \to \mathrm{Mor}(h_Y, h_X)$ *induced by h is bijective.* Prove Yoneda's lemma by showing that one obtains a two-sided inverse to that map by sending $(\vartheta_A)_{A \in \mathrm{Ob}\,\mathcal{C}}$ to $\vartheta_Y(1_Y)$.

(c) Show that h is an anti-equivalence of $\mathcal{C}$ with the full subcategory of $\mathbf{Sets}^{\mathcal{C}}$ consisting of all representable functors $\mathcal{C} \to \mathbf{Sets}$.

7.34. Formulate the analogue of the previous exercise for the contravariant functors $h^X = \mathrm{Mor}_{\mathcal{C}}(-,X)\colon \mathcal{C} \to \mathbf{Sets}$.

7.35. In Section 6 we proved the following: if E and F are two groups such that for all groups G there is a functorial bijection $\mathrm{Hom}(E,G) \to \mathrm{Hom}(F,G)$, then E and F are isomorphic. Deduce this statement from Yoneda's lemma (Exercise 7.33(b)) and Exercises 7.27 and 7.2(b).

7.36. Construct a category $\mathcal{C}$ that has one object and one morphism, and a category $\mathcal{D}$ that has five objects and fifteen morphisms, such that $\mathcal{C}$ and $\mathcal{D}$ are equivalent without being isomorphic.

8. Rings and Modules

We assume that the reader is already familiar with the basic definitions of a ring. We summarize them below.

Definition. *A* **ring** *R is a triple $(R, m, 1)$ where*

(i) R is an additively written group;

(ii) $m : R \times R \to R$, denoted $m(r,s) = rs$, is a bilinear map such that $(rs)t = r(st)$ and $1r = r = r1$ for all $r, s, t \in R$.

Remark. We did not explicitly require R to be an abelian group, although the first condition from the above definition appears at first sight to assume this fact. From previous results about tensor products, we know that the subgroup in R generated by the image of the bilinear map, $m(R \times R) \subset R$, is abelian. However, the map m is surjective by the second axiom, i.e. $1r = r$, and hence $m(R \times R) = R$ and R is abelian.

Definition. *Let A, B be rings. A* **ring homomorphism** *is a map $\phi\colon A \to B$, such that*

(i) ϕ induces a group homomorphism on A and B, viewed as additive groups;

(ii) $\phi(1_A) = \phi(1_B)$;

(iii) $\phi(rs) = \phi(r)\phi(s)$ for all $r, s \in A$.

Definition. *A* **subring** *M of a ring R is an additive subgroup of R such that*

(i) M is closed under the operation of m, i.e. $m(rs) \in M$ for all $r, s \in M$;

(ii) $1 \in M$.

Here are some examples of rings:

(i) **Z**, **R**, **Q**, **C** all form rings under the usual multiplication law.

(ii) We also have the polynomial ring

$$R[x] = \{\sum_{i=0}^{n} a_i x^i : a_i \in R, n \geq 0\}.$$

(iii) The ring $M(n \times n, K)$ of $n \times n$ matrices with entries in a field K is a non-commutative ring if $n > 1$.

(iv) The **zero ring**, where $0 = 1$. This is a terminal object in the category of rings.

(v) Given a group G and ring R, we can form the **group ring** $R[G]$ of formal finite sums $\sum_{x \in G} r_x x$, where only finitely many r_x are non-zero. We define the product

$$(\sum_{x \in G} a_x x)(\sum_{y \in G} b_y y) = \sum_{x \in G}\sum_{y \in G} a_x b_y xy = \sum_{z \in G}(\sum_{xy=z} a_x b_y) z.$$

Note that G is not required to be commutative. The group ring is commutative if and only if the group G is commutative or R is the zero ring.

We now turn our attention to an important class of rings which leads directly into our discussion of modules. Let M be an abelian group. We define the *endomorphisms* of M as the homomorphisms of M into itself, and

$$\mathrm{End}(M) = \mathrm{Hom}(M, M).$$

Since M is an abelian group, $\mathrm{End}(M)$ can be given an abelian group structure by defining addition pointwise, i.e.

$$(f+g)(m) = f(m) + g(m), \forall f, g \in \mathrm{End}(M), m \in M.$$

This sum $f+g$ is still a homomorphism, as is easily checked, and the additive identity is the "zero" homomorphism δ,

$$\delta(m) = 0, \forall m \in M.$$

Moreover, $\mathrm{End}(M)$ forms a ring by defining multiplication as the composition of functions: $m(f,g) = f \circ g$. The multiplicative identity is the identity homomorphism,

$$1_M(m) = m, \forall m \in M.$$

All axioms are easily verified. As an example, consider $\mathrm{End}(C_n) \cong \mathbf{Z}/n\mathbf{Z}$, where C_n is a cyclic group of order n and we identify for each $k \in \mathbf{Z}/n\mathbf{Z}$ the mapping $f_k\colon x \to x^k$, where $x \in C_n$.

Definition. *A* **unit** *in a ring R is an element $a \in R$ such that there exists $b, c \in R$ satisfying $ca = 1, ab = 1$. Note that this immediately implies $c = c(ab) = (ca)b = b$.*

The set of units in a ring is a multiplicative group, as is easily checked, which is often denoted R^* or $U(R)$. Note that an element may or may not be a unit depending on the ring within which one views the element. For example, the element 2 is not a unit in $\mathbf{Z}$, but $2 \in \mathbf{Z} \subset \mathbf{Q}$, and 2 *is* a unit in $\mathbf{Q}$. In general, we have for a subring R of a ring T that $R^* \subset R \cap T^*$. As the above example makes clear, it is not necessarily true that $R^* = R \cap T^*$, although this is true if both R and T are finite rings.

Definition. *Let $x, y \in R$. We call x and y* **zero divisors** *if $x \neq 0$ and $y \neq 0$, but $xy = 0$.*

We also have a notion analogous to the center of a group. We use the same term, although in the case of a ring it refers to the multiplicative operation, under which the ring need not form a group.

Definition. *Let R be a ring. The* **center** *of R is the set of elements*

$$Z(R) = \{a \in R : ar = ra, \forall r \in R\}.$$

Note that the center of a ring is a subring, as is easily checked. Next, we shall discuss the concept of modules. Let R be a ring. Then,

Definition. *A* **left** R**-module** *is an additively written group* M *with a bilinear map* $\beta: R \times M \to M$ *such that for all* $r, s, \in R, m \in M$*, we have*

$$\beta(r,m) = rm, \quad r(sm) = (rs)m \quad \text{and} \quad 1m = m.$$

We can define **right** R**-modules** exactly analogously, using a bilinear map $\beta: R \times M \to M$, written $\beta(r,m) = mr$. The associative law becomes

$$m(rs) = (mr)s, \forall m \in M, r, s \in R.$$

As in the case of rings, we do not explicitly require the additive group M to be abelian. The same argument applies in this case and guarantees that M is indeed abelian. It is important to note, however, that the ring R in the above definition need *not* be commutative.

Remark. For a ring R with multiplication m, the *opposite* ring R^{opp} has the same additive group as R, but its multiplication m^{opp} is defined by $m^{\text{opp}}(r,s) = m(s,r)$. With this definition, a right R-module is the same as a left module over R^{opp}. Namely, given a right R-module with bilinear map β, we can define a bilinear map $\beta': R^{\text{opp}} \times M \to M$ as

$$\beta'(r,x) = r * x = \beta(r,x) = xr,$$

and the first axiom is satisfied by β' since

$$s * (r * x) = (xr)s = x(rs) = x(s * r) = (r * s) * x,$$

where $r * s$ denotes multiplication in R^{opp}.

Recall from our treatment of (left) group actions that an action of G on S naturally gives a homomorphism $\Phi: G \to Perm(S)$, by defining

$$g \mapsto (\Phi(g): x \mapsto gx).$$

Conversely, such a homomorphism immediately yields a group action of G on S. Similarly, we see that any ring homomorphism $\Psi: R \to \text{End}(M)$ induces a left R-module structure on M. Namely, we define the bilinear map $\beta: R \times M \to M$ such that $(r,m) \mapsto \Psi(r)(m) = rm$. Then we have

$$\beta(r_1 + r_2, m) = \Psi(r_1 + r_2)(m) = \Psi(r_1)(m) + \Psi(r_2)(m) = \beta(r_1, m) + \beta(r_2, m),$$

since Ψ is a homomorphism into $\text{End}(M)$. Also, we have

$$\beta(r, m_1 + m_2) = \Psi(r)(m_1 + m_2) = \Psi(r)(m_1) + \Psi(r)(m_2) = \beta(r, m_1) + \beta(r, m_2),$$

since $\Psi \in \mathrm{End}(M)$. Hence the map is bilinear. The associativity follows from the fact that multiplication in $\mathrm{End}(M)$ is the composition of maps, and that Ψ is a homomorphism:

$$r(sm) = \Psi(r)(\Psi(s)(m)) = \Psi(rs)(m) = (rs)m.$$

The action of the identity follows since $1_R m = \Psi(1_R)(m) = 1_M(m) = m$, i.e. Ψ carries the identity to the identity.

Similarly, any R-module yields a homomorphism Ψ from the ring R to $\mathrm{End}(M)$, by defining $\Psi(r) = (m \mapsto rm)$. Hence a homomorphism Ψ as above is equivalent to giving an R-module structure on M.

Just as sets have subsets, and groups have subgroups, R-modules have submodules:

Definition. *A* **submodule** *N of an R-module M is an additive subgroup of M such that for all $r \in R$ one has $rN \subset N$.*

We may also examine the class of R-modules as a category (or rather as the objects of a category). The morphisms in this category would be defined as follows:

Definition. *A* **morphism of R-modules** *$f \colon M \to N$ is a group homomorphism with respect to the additive groups, such that for all $r \in R$ and $m \in M$, we have $r(f(m)) = f(rm)$. This property is known as* **R-linearity**.

The notions of direct sum and direct product can also be defined for modules. Let $(M_i)_{i \in I}$ be a family of R-modules. We define their *direct product* and *direct sum* to be

$$\prod_{i \in I} M_i \quad \text{and} \quad \bigoplus_{i \in I} M_i,$$

respectively the direct sum and product of their abelian groups, with coordinatewise multiplication completing the R-module structure. That is, if $(m_i)_{i \in I}$ is an element in the direct product (or sum) of the M_i, then for all $r \in R$ one defines $r \cdot (m_i)_{i \in I}$ to be $(rm_i)_{i \in I}$. This definition is consistent with the general categorical definition of direct product (or sum), as the reader can easily verify.

It is instructive to visit a few examples of R-modules to familiarize ourselves with the concept. Here are the examples:

(i) Any (left) ideal of R is an R-module. One can view R itself as an R-module by using the multiplication operation on R as the bilinear map. In this case, the left ideals are the same as the R-submodules. In fact, two sided ideals (such as R itself) are both left and right R-modules.

(ii) A $\mathbf{Z}$-module is the same as an abelian group. In other words, given any abelian group, there is exactly one way to impose a $\mathbf{Z}$-module structure on it.

(iii) Let K be a field. A K-module is the same as a vector space over K. Here, the equivalence follows from a close reading of the definition of a vector space over K, which can be seen to be just a rewording of the definition of a K-module.

(iv) Let K be a field and $K[x]$ the ring of polynomials in x with coefficients in K. A $K[x]$-module is the same as a vector space V_K over K, together with a K-linear map $\epsilon\colon V_K \to V_K$. Note that a K-linear map is the same as our standard notion of a linear map between vector spaces.

(v) Let $K[G]$ be a group ring. A $K[G]$-module is the same as a vector space V_K over K, together with a linear representation of G on V_K, that is a homomorphism from G to $\mathrm{Aut}_K(V_K)$.

We noted in the above example that a **Z**-module is the same as an abelian group. In fact, much of the basic theory of modules carries over from the basic theory of abelian groups. We itemize some of the direct analogues here, just to emphasize the similarities.

(i) We have already defined the notion of an R-submodule, which is an additive subgroup stable under the action of R.

(ii) We have the notion of a *factor module* M/N. This is just the quotient group with the natural induced R-action.

(iii) We have already defined a *homomorphism* or R-modules; these are R-linear group homomorphisms $f\colon M \to N$.

(iv) We have the basic isomorphism theorem: Given a morphism f as above:

$$M/\ker(f) \cong f(M) \subset N,$$

the homomorphic image $f(M)$ being a submodule of N.

(v) The theorem of Jordan-Hölder also carries over, and in fact simplifies a little, since we do not need to worry about the normality of subgroups. Hence we have the notions of *composition series* of modules, and of *simple modules.* (Simple modules have exactly two submodules, namely, 0 and M itself.)

(vi) We also have exact sequences of modules and the Snake Lemma.

(vii) We can also form *direct products* and *direct sums* of modules, as shown previously.

We have a characterization of simple R-modules.

Proposition. *The following statements are equivalent:*

(i) M *is a simple* R*-module;*

(ii) $M \cong_R R/I$*, where* I *is a maximal left ideal in* R*.*

Proof. We give a sketch of a proof. There is a bijective correspondence between the set of left ideals J in R such that $I \subset J \subset R$, and submodules of R/I, by the map $J \mapsto J/I$. So a module of the form R/I, is simple if and only if I is maximal.

Moreover, we can show that any simple module is of this form. Pick any non-zero element $x \in M$, which exists since M is nonzero, and define the map $f\colon R \to M$ by $r \mapsto rx$. This is R-linear, so $f(R) \subset M$ is a non-zero submodule of M, and since M is simple, we have $f(R) = M$. So $M \cong_R R/\ker(f)$.

Remark. One can ask whether there exists a *unique* maximal ideal such that $M \cong_R R/I$. Since the choice of x in the above proof was arbitrary, a priori this is not clear. It turns out that the ideal is unique if R is commutative, but in the non-commutative case it may be wrong.

We can also approach the concept of the *free module* by taking a step back into category theory. We have many categories (**Group**, the category of groups; **Ab**, the category of abelian groups; ${}_R$**Module**, the category of R-modules) which may be mapped via an appropriate forgetful functor into **Set**, the category of sets. Let us examine the case of the category of groups, with which we are already familiar. We have the "free" functor F which takes a set S to $F(S)$, the free group on S, and is a left adjoint to the forgetful functor f:

$$\mathbf{Group} \underset{f}{\overset{F}{\rightleftarrows}} \mathbf{Set}.$$

In other words, given a group G, and a set S, there is a natural bijection

$$\mathrm{Hom}_{\mathbf{Group}}(F(S), G) \leftrightarrow \mathrm{Hom}_{\mathbf{Set}}(S, f(G)).$$

We are already aware of the natural analogous "free" functor for the case of the category of abelian groups. We now consider the category of R-modules. Recall that any abelian group can be viewed as a $\mathbf{Z}$-module. Thus, for any set S, the functor $F\colon \mathbf{Set} \to \mathbf{Ab}$ maps

$$S \mapsto \bigoplus_{i \in S} \mathbf{Z}.$$

The extrapolation to modules over a general ring R is automatic. We define the *free R-module* on a set S, denoted (as usual) by $F(S)$, by

$$F(S) = \bigoplus_{i \in S} R.$$

The map F is a covariant functor from **Set** to ${}_R$**Module**, as the reader can verify.

When talking about the direct product (or sum) of a set of identical modules, one often uses an alternative notation. In the case of a direct product, one often writes

$$R^S = \prod_{i \in S} R,$$

and for a direct sum, one often writes

$$R^{(S)} = \bigoplus_{i \in S} R.$$

So if ${}_R\operatorname{Hom}$ denotes the set of R-linear homomorphisms between two R-modules, we again have

$${}_R\operatorname{Hom}(R^{(S)}, M) \cong \operatorname{Maps}(S, M),$$

showing that the free functor is again a left adjoint of the forgetful functor. Free modules will be discussed in greater depth later in the notes.

Some problems may be approached from a module-theoretic perspective in order to gain insight. If two rings R and R' share certain characteristics, then the classes of R-modules and R'-modules may also show certain similarities. For example, the rings $\mathbf{Z}$ and $K[x]$ share the characteristic of being principal ideal domains. The $\mathbf{Z}$-modules are abelian groups, and we have a well-known result classifying the finite abelian groups, namely that every finite abelian group is isomorphic to a direct product of finite cyclic groups. We can attempt to make a similar classification of $K[x]$-modules, which are equivalent to vector spaces V_K over K equipped with a K-linear endomorphism ϵ on V_K. That is, given a vector space V_K and a K-linear endomorphism ϵ on V_K, we can view this as a $K[x]$-module in the following way. Given a polynomial $f(x) = \sum_i a_i x^i$ in $K[x]$ and an element v in V_K, we define a bilinear map β as follows: $\beta(f(x), v) = \sum_i a_i \epsilon^i v$. It is easy to check that this defines a $K[x]$-module structure on V_K. The reader can also check that the process reverses naturally; namely, given a $K[x]$-module M, one can view it as a vector space over K equipped with a K-linear endomorphism.

Since we are restricting our attention to finite abelian groups in the classification, we must make a similar restriction for the vector spaces. In fact, the restriction is the intuitive one, namely that we restrict our attention to the finite-dimensional vector spaces. The classification is found by examining the result in a module-theoretic form. If A is a finite $\mathbf{Z}$-module, then

$$A \cong \bigoplus_{i=1}^{t} \mathbf{Z}/p_i^{r_i}\mathbf{Z},$$

where p_i is a prime in the ring $\mathbf{Z}$. There is a completely analogous statement which can be made for finite-dimensional vector spaces over K with a K-linear endomorphism ϵ. This analogy is most apparent when K is algebraically closed. If $\langle V_K, \epsilon \rangle$ is our n-dimensional $K[x]$-module, we have

$$\langle V_K, \epsilon \rangle \cong \bigoplus_{i=1}^{t} K[x]/((x - \lambda_i)^{r_i} \cdot K[x]), \quad \sum_{i \in I} r_i = n$$

since $(x - \lambda_i)$ are the prime polynomials in the ring $K[x]$. The linear endomorphism on this vector space sum is the $n \times n$ matrix:

$$\epsilon = \begin{bmatrix} \underbrace{\begin{matrix} \lambda_1 & 1 & 0 & \cdots & \\ 0 & \lambda_1 & 1 & \ddots & \\ \vdots & \ddots & \ddots & \ddots & \\ & & & \lambda_1 & 1 \\ & & & & \lambda_1 \end{matrix}}_{r_1 \text{ columns}} & & & 0 \\ & \underbrace{\begin{matrix} \lambda_2 & 1 & \\ & \ddots & \ddots \\ & & \lambda_2 \end{matrix}}_{r_2 \text{ columns}} & & \\ 0 & & \ddots & \\ & & & \underbrace{\begin{matrix} \lambda_t & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_t & 1 \\ & & & \lambda_t \end{matrix}}_{r_t \text{ columns}} \end{bmatrix}$$

In fact the proof for the classification of finite abelian groups can be translated directly to prove this classification of finite-dimensional vector spaces over K with a K-linear endomorphism ϵ. This illustrates the usefulness of approaching certain objects as R-modules.

For the next section, let R be a commutative ring.

Definition. *An* **R-algebra** *is a ring A with a ring homomorphism*

$$f\colon R \to A$$

such that the image of R under f lies inside the center of the ring A, i.e. $f(R) \subseteq Z(A)$. A morphism of R-algebras is a ring homomorphism $g : A \to B$ such that the following diagram is commutative:

$$\begin{tikzcd} & R \arrow[dl, "f_A"'] \arrow[dr, "f_b"] & \\ Z(A) \arrow[d] & & Z(B) \arrow[d] \\ A \arrow[rr, "g"] & & B \end{tikzcd}$$

First, we note that the ring A in the above definition need not be commutative. Also, in many cases with which we are concerned, the map $f\colon R \to A$ is injective; hence we can view the R-algebra A as a ring with R lying inside its center as a subring. This latter fact serves as an excuse for a notational shorthand in which the ring R is identified with its image $f(R)$, and the map is $f\colon r \mapsto f(r) = r \in A$. This provokes confusion in cases where f is not injective, since it makes it necessary to distinguish the condition "$r = s$ in R" from the condition "$r = s$ in A." Consider, for example, the case where $\mathbf{Z}/n\mathbf{Z}$ is viewed as a $\mathbf{Z}$-algebra. The notation, however, is here to stay.

We note here that when the ring R happens to be a field (and A is non-zero), the ring homomorphism $f\colon R \to A$ is injective, and therefore R can be canonically identified with its image in A. Thus, for K a field, a non-zero K-algebra *is*, effectively, a ring containing K as a subring.

It is also useful to note that if A is an R-algebra, then any factor ring A/M of A can also be given an R-algebra structure in a natural way. Namely, we can compose the ring homomorphism f with the projection map:

$$R \xrightarrow{f} A \xrightarrow{\pi} A/M.$$

The composite ring homomorphism $\pi \circ f$ is the required ring homomorphism from R to A/M. We now present some examples.

(i) Any ring R can be viewed as a $\mathbf{Z}$-algebra. Since $\mathbf{Z}$ is an initial object in the category of rings, there exists only one homomorphism $f\colon \mathbf{Z} \to R$, and hence the category of $\mathbf{Z}$-algebras and the category of rings are equivalent.

(ii) Let K be an extension field of the rationals $\mathbf{Q}$. Then K is a $\mathbf{Q}$-algebra, where the homomorphism $f\colon \mathbf{Q} \to K$ is simply inclusion.

(iii) Given a group G and a ring R, the group ring $R[G]$ forms an R-algebra, where again the homomorphism is simply inclusion.

The additive group of an R-algebra A can be given an R-module structure in the natural way by defining the bilinear map

$$\begin{array}{rccc} \beta\colon & R \times A & \to & A \\ \beta\colon & (r,a) & \mapsto & r * a = f(r)a. \end{array}$$

This map is bilinear, since f is a homomorphism and because multiplication in A is bilinear. Namely,

$$(r_1 + r_2) * a = f(r_1 + r_2)a = (f(r_1) + f(r_2))a = f(r_1)a + f(r_2)a = r_1 * a + r_2 * a,$$

and similarly

$$r * (a_1 + a_2) = f(r)(a_1 + a_2) = f(r)a_1 + f(r)a_2 = r * a_1 + r * a_2.$$

Moreover, the action of the identity is satisfied since f carries the identity to the identity,

$$1_R * a = f(1_R)a = 1_A a = a,$$

for all $a \in A$. Associativity follows from associativity of multiplication in A. Hence all axioms are satisfied and A is indeed an R-module.

When we examine a homomorphism between groups, we are often interested in its kernel and image. The same is true for ring homomorphisms, but the analogies are not entirely trivial. In the case of a group homomorphism, its kernel is a normal subgroup and its image is isomorphic to a quotient group. However we have no notion of a "normal subring" and in fact the kernel of a homomorphism is rarely a subring at all. It is an ideal.

Definition. *Let R be a ring. A* **left ideal** *I is an additive subgroup of R such that for all $r \in R$ one has $rI \subset I$.*

A *right ideal* is defined as one would expect, and a *two-sided ideal* of R is defined to be an additive subgroup that is both a left and right ideal of R.

It is worthwhile to emphasize that a left ideal I is a submodule of R, if we view R as a (left) R-module. Many of the results we will prove about modules will, in particular, be relevant to the study of ideals.

From the definition of an ideal, we can immediately present the following theorem.

Theorem. *Let R and T be rings, with a ring homomorphism $\Phi\colon R \to T$. The kernel of Φ is a two-sided ideal of R.*

Proof. If k_1 and k_2 are in the kernel of Φ, then

$$\Phi(k_1 + k_2) = \Phi(k_1) + \Phi(k_2) = 0 + 0 = 0$$

so $k_1 + k_2$ is in the kernel of Φ. Furthermore, given any $r \in R$, and k in the kernel of Φ,

$$\Phi(rk) = \Phi(r)\Phi(k) = \Phi(r) \cdot 0 = 0 \quad \text{and} \quad \Phi(kr) = \Phi(k)\Phi(r) = 0 \cdot \Phi(r) = 0$$

so rk and kr are in the kernel of Φ. This shows that the kernel of Φ is a two-sided ideal.

An ideal being two-sided corresponds to a subgroup being normal. Given a ring and a two-sided ideal, the notion of a quotient ring presents itself immediately.

Definition. *Given a ring R and a two-sided ideal I of R, the* **quotient ring**, *R/I is defined to be the quotient group of the additive groups, with multiplication defined canonically. That is, given $r_1 + I$ and $r_2 + I$ in R/I, their multiplicative product is defined to be $r_1 r_2 + I$.*

That this multiplication is well-defined and satisfies the ring axioms is left as an exercise for the reader. Furthermore, the reader may wish to check what happens when I is not required to be a two-sided ideal.

So if I is the kernel of a homomorphism $\Phi\colon R \to T$, the image of Φ is isomorphic to the quotient ring, R/I.

It was mentioned that the kernel of a homomorphism is *rarely* a subring. In any ring R, there is only one subset which is simultaneously an ideal and a subring.

Theorem. *Let R be a ring, and I a left ideal of R that contains a unit. Then $I = R$.*

Proof. Let u be a unit in I. Let v be such that $vu = 1$. Then for any r in R, rv is also in R, so $(rv)u \in I$. But $(rv)u = r(vu) = r(1) = r$. Therefore $I = R$.

Since every subring contains the identity (which is certainly a unit), the only ideal of R which is a subring is the ring R itself.

Certain ideals are of particular interest. A left ideal I of R is called a *principal (left) ideal* if there is some $r \in R$ such that $Rr = I$. In a commutative ring R, a proper ideal P is called a *prime ideal* if for every a and b not in P, the product ab is also not in P, and an ideal M is called a *maximal ideal* if for every ideal I, such that $M \subset I \subset R$, either $I = M$ or $I = R$ but not both.

This last type of ideal has set-theoretic interest as well. There is a proof of the existence of maximal ideals in any commutative ring, but it requires the axiom of choice. The details of the proof will not be discussed, other than that it follows fairly straightforwardly from the Zorn's Lemma formulation of the axiom of choice.

Fact (assuming Axiom of Choice). *Let R be a non-trivial commutative ring. Then there is a maximal ideal M of R.*

From this we can prove a slightly stronger (and more useful) statement about maximal ideals.

Theorem. *Let R be a commutative ring, and let I be a proper ideal of R. Then there is a maximal ideal M of R such that $I \subset M$.*

Proof. Since R is commutative, I is automatically two-sided. Therefore the quotient ring R/I is also commutative (and since I is proper, it is non-trivial). By the above fact, this ring contains a maximal ideal, M'. If Φ is the quotient map from R to R/I, let M be the inverse image $\Phi^{-1}(M')$. Since Φ is a ring homomorphism, a simple check shows that M is an ideal. Furthermore, for any ideal N of R, if $M \subset N \subset R$, then $M' \subset \phi(N) \subset R/I$. This shows that M is maximal, and since $0 \in M'$, we have that $I \subset M$.

The theorem has the following corollary:

Corollary. *Let R be a commutative ring, and let $\langle M_i \rangle$ be the set of maximal ideals of R. Then*

$$\bigcup_i M_i = R \setminus R^*.$$

Proof. Suppose $x \in R \setminus R^*$. Then the principal ideal Rx does not contain the identity, and is therefore proper. So there is a maximal ideal M which contains x.

In the other direction, suppose $x \in M$, a maximal ideal. Then since M can not contain any units, x is not a unit. Therefore $x \in R \setminus R^*$.

Definition. *A* **field** *is a commutative ring R with the property $R^* = R \backslash \{0\}$.*

A ring that satisfies the above property but is not necessarily commutative is called a *skew field.* Note that the above set-theoretic equality automatically requires that $0 \neq 1$, since 1 is a unit in R; so the zero ring is not a field. Here are two examples of fields.

(i) $\mathbf{Q}$, the rational numbers, form a field.
(ii) $\mathbf{Z}/p\mathbf{Z}$, for $p \in \mathbf{Z}$, forms a field if and only if p is prime. This also can be taken as the definition of a prime in the ring of integers; note also that this requires that we do not view 1 as a prime number.

Theorem. *Let R be a commutative ring. Then R is a field if and only if R has exactly two ideals, namely, the zero ideal and itself.*

Proof. Suppose R has exactly two ideals. Let $a \in R$ be a non-zero element and consider the principal ideal Ra. Since this is non-zero, $Ra = R$. In particular, there exists an element $b \in R$ with $ba = 1$. Hence a is a unit.

Now suppose R is a field. Let I be a non-zero ideal. Then it contains some non-zero element $a \in I$. Since a is a unit, by an earlier theorem concerning ideals with units, we have $I = R$. So any non-zero ideal is all of R, and R has exactly two ideals.

Note that the conditions in the theorem above exclude that R is the zero ring. Given this characterization of fields for commutative rings R, we can easily show that given any ring R and a maximal ideal I, the quotient ring R/I is a field. Before the proof, we mention two facts which are easy to check directly from the axioms. Namely, if $f\colon A \to B$ a ring homomorphism, A, B commutative rings, and an ideal $I \subset B$, then $f^{-1}(I) \subset A$ is also an ideal. Also, if A, B are as above, $f : A \to B$ a surjective ring homomorphism and I an ideal in A, then $f(I)$ is an ideal.

Theorem. *Let R be a commutative ring. Then an ideal I of R is maximal if and only if R/I is a field.*

Proof. Let I be an ideal of R such that R/I is not a field. Then there must exist some non-trivial proper ideal J' of R/I. Let $\Phi\colon R \to R/I$ be the quotient homomorphism. Let J be the ideal $\Phi^{-1}(J') \subset R$. Then $I \subset J \subset R$, and these inclusions are strict since J' is a proper ideal of R/I. Hence I is not maximal.

In the other direction, let I be an ideal of R such that R/I is a field. Let J be an ideal such that $I \subset J \subset R$. Then $\Phi(J)$ is an ideal in R/I, and therefore must be 0 or R/I. This means that either $J = R$ or $J = I$. This shows that I is maximal.

We now turn our attention to a special class of rings which can naturally be viewed as subrings of fields.

Definition. *A* **domain** *is a non-trivial commutative ring which contains no zero divisors, i.e. has the property that if $ab = 0$, then $a = 0$ or $b = 0$.*

In domains, the following useful fact holds: if $ab = ac$, $a \neq 0$, then $b = c$. This is often called the cancellation law. That any subring of a field is a domain is clear, but in fact given any domain D, there is a field K, and a subring $R \subset K$ such that $R \cong D$. This field K is the field of fractions of D:

We construct the *field of fractions* $Q(R)$ of the domain R. As a set, $Q(R)$ is the quotient of the set $\{(a, b) : a, b \in R, b \neq 0\}$ by the equivalence relation

$$(a_1, b_1) \equiv (a_2, b_2) \Leftrightarrow a_1 b_2 = b_1 a_2 \in R.$$

(That this is an equivalence relation requires the use of the cancellation law in R.) Denote the equivalence class containing a pair (a, b) by $\frac{a}{b}$. Multiplication is defined by

$$\frac{a_1}{b_1} \cdot \frac{a_2}{b_2} = \frac{a_1 a_2}{b_1 b_2},$$

and addition by

$$\frac{a_1}{b_1} + \frac{a_2}{b_2} = \frac{a_1 b_2 + a_2 b_1}{b_1 b_2}.$$

It is straightforward to check that these operations are well-defined on the equivalence classes, and that the ring axioms are satisfied. Moreover, $Q(R)$ is a field, since every non-zero element $\frac{a}{b}$ has an inverse, namely $\frac{b}{a}$.

We remind the reader that a prime ideal P of a commutative ring R is a proper ideal such that for all a, b not in P, then ab is also not in P. The relation between fields and maximal ideals is analogous to the relation between domains and prime ideals. There are other important points to make about prime ideals, but first we need a definition.

Definition. *Let R be a commutative ring. A subset S of R is called a* **multiplicative subset** *if $1 \in S$ and S is closed under multiplication.*

The following statements are all equivalent to P being a prime ideal:

(i) P is an ideal of R whose complement is a multiplicative subset S.
(ii) P is an ideal such that R/P is a domain.
(iii) P is the kernel of a ring homomorphism Φ from R to a field K.

The first statement follows directly from the definitions. The third is equivalent to the second, since R/P is a domain if and only if R/P is a subring of a field. We prove the second statement as follows.

Proof. Suppose that R/P is a domain, and let Φ be the quotient map from R to R/P. If a and b are not in P, then $\Phi(a)$ and $\Phi(b)$ are not 0. Since R/P is a domain, $\Phi(ab) = \Phi(a)\Phi(b)$ is also not zero. Therefore ab is not in P. This shows that P is a prime ideal.

Suppose P is a prime ideal. If $\Phi(a)$ and $\Phi(b)$ are non-zero in R/P, then a and b are not in P. So ab is not in P, and therefore $\Phi(a)\Phi(b) = \Phi(ab)$ is also not zero. This shows that R/P is a domain.

Prime ideals give, other than maximal ideals, rise to an interesting contravariant functor from commutative rings to sets.

Definition. *Let* **CommRing** *be the category of commutative rings and* **Set** *be the category of sets. The functor* Spec: **CommRing** $\to$ **Set** *maps a ring to the set of its prime ideals. Furthermore, if* $\Phi\colon R_1 \to R_2$ *is a ring homomorphism,* Spec *maps* Φ *to the function* $\Phi^*\colon \mathrm{Spec}(R_2) \to \mathrm{Spec}(R_1)$ *where for any ideal* $I \subset R_2$, Φ^* *maps* I *to* $\Phi^{-1}(I)$.

The proof that Spec is a contravariant functor boils down to seeing that given any ring homomorphism Φ, the inverse image of a prime ideal is a prime ideal. The proof for this is fairly straightforward.

Proof. Let $\Phi\colon R_1 \to R_2$ be a ring homomorphism, and let I_2 be a prime ideal of R_2. Let Ψ be a ring homomorphism from R_2 to a field K such that I_2 is the kernel of Ψ. Then the composite $\Psi \circ \Phi$ is a ring homomorphism from R_1 to K. Let I_1 be the inverse image $\Phi^{-1}(I_2)$. Then I_1 is the kernel of $\Psi \circ \Phi$, so I_1 is a prime ideal of R_1.

If S is a multiplicative subset of a ring R, one can construct an R-algebra, denoted $S^{-1}R$, and often called the *ring of fractions of R by S* or the *quotient ring of R by S*, as follows. We define $S^{-1}R$ informally to be the ring of equivalence classes of fractions $\frac{r}{s}$, with addition and multiplication defined as one would expect, but with the equivalence of fractions extended so that $\frac{r}{s} \equiv \frac{r'}{s'}$ if there exists some $\sigma \in S$ such that

$$\sigma s' r = \sigma s r'.$$

Note that this equivalence relation is stronger than the usual equivalence for fractions which requires σ to be 1, and therefore the inclusion map from R to $S^{-1}R$ sending $r \mapsto \frac{r}{1}$ is a ring homomorphism, but not in general injective. Indeed, if S contains 0, $S^{-1}R$ is trivial. However, it should be noticed that there is a natural R-algebra structure on this ring via the above inclusion map.

Theorem. *Let R be a commutative ring, and let S be a multiplicative subset of R. We have that $0 \notin S$ if and only if there exists a prime ideal P in R such that P and S are disjoint.*

Proof. If there exists such a prime ideal P, then since any ideal contains 0, we have $0 \notin S$. Now suppose $0 \notin S$. Consider the set of ideals I of R such that $I \cap S = \emptyset$. Let P be a maximal element of S (which exists by Zorn's lemma). We claim that P is prime. Let $a, b \in R, ab \in P$. Suppose $a \notin P, b \notin P$. Then the intersection of the ideal $P+Ra$ generated by P and a with S is nonempty, i.e. $(P+Ra) \cap S \neq \emptyset$. Similarly, $(P+Rb) \cap S \neq \emptyset$. Thus we have elements $p_1 + r_1 a \in S$, $p_2 + r_2 b \in S$, for some p_1, $p_2 \in P$, r_1, $r_2 \in R$. Since S is a multiplicative subset, we have

$$(p_1 + r_1 a)(p_2 + r_2 b) = p_1 p_2 + p_1 r_2 b + r_1 a p_2 + r_1 r_2 ab \in S.$$

Since each term on the right hand side is contained in P, the sum is also contained in P; but P and S are disjoint. Hence P is prime, as desired.

Definition. *An element r of a ring is* **nilpotent** *if there exists some positive integer n such that $r^n = 0$. The set of nilpotent elements of R is known as the* **nilradical** *of R.*

In the case of a commutative ring R, it can be easily verified that the nilradical of the ring is an ideal. This is not necessarily the case for a non-commutative ring. The nilradical is closely related to the set of prime ideals, as in the following theorem.

Theorem. *Let R be a commutative ring. The nilradical of R is equal to the intersection*

$$\bigcap_{\mathrm{Spec}(R)} P.$$

Proof. If x is nilpotent, for any prime ideal P, there exists a positive integer n such that

$$\overbrace{x \cdot x \cdots x}^{n\text{ times}} = 0 \in P.$$

By definition of a prime ideal, this implies that either $x^{n-1} \in P$ or $x \in P$ (or both). By induction on n, we see that $x \in P$.

In the other direction, if x is not nilpotent, let S_x be the subset

$$S_x = \{x^n : n \in \mathbf{Z}, n > 0\}.$$

Here we interpret x^0 to be 1. Thus we see that S_x is a multiplicative subset. Since x is not nilpotent, S_x does not contain 0. Therefore, by the previous theorem, there is a prime ideal that is disjoint from S_x, and so in particular does not contain x. Then x is not in the intersection of the prime ideals.

We drop for the moment our usual condition that R is commutative.

Let R_i, $i \in I$, be a family of rings. Then $R = \prod_{i \in I} R_i$ is naturally a ring by componentwise operations. In this section, we restrict our attention to cases where I is finite. For concreteness we take the case of a product of two rings; the other cases follow by induction.

Let $R = R_1 \times R_2$. Then R contains two elements, namely $(1,0)$ and $(0,1)$, such that $(1,0)^2 = (1,0)$, and $(0,1)^2 = (0,1)$. This property is important enough to formally name it, and will come in handy when we consider the decomposition of rings into direct products.

Definition. *Let R be any ring. An element $e \in R$ is called an* **idempotent** *if $e^2 = e$.*

Many questions about the product ring R reduce to questions about the rings R_1 and R_2. For instance, the group of units R^* of R consist of precisely those elements both of whose components are invertible. Namely, one easily checks that $R^* = R_1^* \times R_2^*$.

Similarly, any left ideal I of R can be uniquely written in the form $I_1 \times I_2$, where each I_i is a (left) ideal in R_i. This is a consequence of a fact proven below about R-modules, since any ideal $I \subset R$ is an R-module.

For a two-sided ideal $I = I_1 \times I_2$, we have the convenient characterization of the structure of the quotient ring R/I, namely

$$R/I \cong (R_1/I_1) \times (R_2/I_2),$$

by the natural map $\Phi\colon (a,b) + (I_1 \times I_2) \mapsto (a + I_1, b + I_2)$. It is straightforward to check that this is well-defined and an isomorphism.

In the case when R_1, R_2 are both commutative, R is also commutative. Since we know that I is prime or maximal if and only if R/I is a domain or a field, respectively, and we have the above characterization of R/I, we see that I is a prime ideal if and only if $I = I_1 \times I_2$ where one of I_1 or I_2 is the whole ring, and the other is prime. The analogous statement holds for maximal ideals.

Proposition. *An ideal $I = I_1 \times I_2$ in the product ring $R = R_1 \times R_2$ is prime (maximal) if and only if one of I_1, I_2 is equal to the corresponding ring R_1, R_2, and the other is prime (maximal).*

Proof. Let I be prime in R. Consider the elements in the quotient ring $(0,1)$ and $(1,0)$. We know that $(0,1)(1,0) = (0,0) = 0 \in R/I$. Since R/I is a domain, this means that one of $(1,0)$ or $(0,1)$ must be 0 in R/I. This immediately implies that one of $1_{R_1} \in I_1$ or $1_{R_2} \in I_2$, i.e. one of the following holds: $I_1 = R_1$ or $I_2 = R_2$. Without loss of generality, let $R/I \cong R_1/I_1$. Clearly I is prime (maximal) if and only if I_1 is prime (maximal).

As an extension of the above statement, we also know that if R_1 and R_2 are commutative, the spec of the ring R is the disjoint union of those of the rings R_1 and R_2, namely

$$\operatorname{Spec}(R) = \operatorname{Spec}(R_1) \amalg \operatorname{Spec}(R_2).$$

The spec of a ring can be endowed with a certain natural topology (which we will not define here), and it turns out that if the spec of a product ring is decomposed as above, its topology is the canonical topology on the disjoint union of the topological spaces $\operatorname{Spec}(R_1)$ and $\operatorname{Spec}(R_2)$. In particular, if R_1 and R_2 are non-trivial, $\operatorname{Spec}(R)$ is not connected. Conversely, if R is *not* a product of non-trivial rings, $\operatorname{Spec}(R)$ *is* connected. Hence (by abuse of language) we shall call a commutative ring *connected* if it cannot be decomposed as a direct product of nontrivial rings.

Much of what we have said applies also to R-modules. Let M_1 be an R_1-module, and let M_2 be an R_2-module. The direct product $M = M_1 \times M_2$ of abelian groups naturally has an R-module structure, where R is the direct product $R = R_1 \times R_2$.

The surprising fact is that *any* R-module is of the above form, namely any R-module M can be uniquely decomposed into submodules M_1, M_2 such that $M = M_1 \times M_2$.

Theorem. *Any $R_1 \times R_2$-module M can be uniquely decomposed into a direct product of modules, i.e. $M = M_1 \times M_2$ where M_i is an R_i-module, $i = 1, 2$.*

Proof. Let $M_1 = (1,0)M$, $M_2 = (0,1)M$. These are clearly R-submodules of M, since $(1,0)$ and $(0,1)$ lie in the center of R. Now consider $M_1 \cap M_2$. Since $(1,0), (0,1)$ are both idempotents, $(1,0)(1,0)m = (1,0)m, \forall m \in M$, so $(1,0)$ acts as the identity on M_1, and acts as 0 on M_2. Hence $M_1 \cap M_2 = 0$. Since for any $m \in M$, we have $m = (1,1)m = (1,0)m + (0,1)m$, we have $M = M_1 \oplus M_2$. But the M_i are naturally R_i-modules, by the action $r_1 m_1 = (r_1, 0)m_1$ (similarly for R_2).

An analogous statement can be made for morphisms of $(R_1 \times R_2)$-modules. Certainly, if M_i, N_i are R_i-modules ($i = 1, 2$), and $f_i \colon M_i \to N_i$ are R_i-linear maps, then we can define a R-linear map $f \colon M \to N$ componentwise, namely $f = (f_1, f_2)$. We can show that every map f arises in this way; i.e. every R-linear map can be uniquely decomposed as above. Since N is a direct product of abelian groups N_1, N_2, we already have $f_i \pi_i = \pi_i f$, the projections onto the components, which are immediately homomorphisms of abelian groups. We only need to show that each f_i is R_i-linear. But

$$\begin{aligned}
(r_1, 0)\pi_1 f(x,y) &= (r_1, 0)(f_1(x,y), 0) \\
&= (r_1 f_1(x,y), 0) \\
&= \pi_1((r_1, 0) f(x,y)) \\
&= \pi_1 f((r_1, 0)(x,y)),
\end{aligned}$$

since f is R-linear. Similarly, f_2 is R_2-linear. Hence we have $f = (f_1, f_2)$ where the f_i are R_i-linear, and hence f is an R-module homomorphism.

Another way of phrasing this in categorical language, is to say that if ${}_R\mathbf{Module}$ is the category of R-modules, then for all rings R_1 and R_2, we have an equivalence of categories between ${}_{(R_1 \times R_2)}\mathbf{Module}$ and ${}_{R_1}\mathbf{Module} \times {}_{R_2}\mathbf{Module}$, the direct product of categories being defined in the obvious way.

We can apply certain operations to ideals to generate more ideals. Intersection, addition, and multiplication are all well-defined on ideals as follows:

Let R be a ring, and let I and J be additive subgroups of R.

(i) The intersection $I \cap J$ is a left (right) ideal if both I and J are left (right) ideals.

(ii) The sum $I + J = \{x + y : x \in I \text{ and } y \in J\}$ is a left (right) ideal if both I and J are left (right) ideals.

(iii) The product $I \cdot J = \{\sum_{i=1}^{t} x_i y_i : t \in \mathfrak{N}, x_i \in I, y_i \in J\}$ is a left (right) ideal if I is a left ideal (and J is any subset of R).

When I and J are two-sided, we have

$$I \cdot J \subset I \cap J \subset I + J,$$

as the reader can easily verify. The simplest examples to illustrate these ideals come from taking the ring R to be $\mathbf{Z}$. Let $n\mathbf{Z}$ and $m\mathbf{Z}$ be two ideals in $\mathbf{Z}$.

Examples.

(i) The sum $n\mathbf{Z} + m\mathbf{Z}$ is the ideal $t\mathbf{Z}$, where t is the greatest common divisor of n and m.

(ii) The intersection $n\mathbf{Z} \cap m\mathbf{Z}$ is the ideal $t\mathbf{Z}$, where t is the least common multiple of n and m.

(iii) The product $n\mathbf{Z} \cdot m\mathbf{Z}$ is the ideal $t\mathbf{Z}$ where $t = nm$.

We now proceed to a ring-theoretic generalization of a well-known elementary number theory result. In this section, we assume R to be commutative. Most statements can, with sufficient care, be made for general rings, but we do not bother to do so here.

Definition. *Two ideals I, J of R are* **coprime** *if $I + J = R$.*

Remark. Equivalently, we say I, J are coprime if there exists $x \in I, y \in J$ such that $x + y = 1$. The equivalence is easy to see: if such x and y exist, then for any $r \in R$, we have $r = r1 = r(x + y) = rx + ry \in I + J$. On the other hand, if $I + J = R$, since $1 \in R$ there must exist such x and y. Furthermore it is worth noting that any two distinct maximal ideals must be coprime. (The reader can verify this easily.)

Chinese Remainder Theorem. *Let R be a commutative ring, I, J coprime ideals in R. Then $IJ = I \cap J$, and the ring homomorphism $f: R \to (R/I) \times (R/J)$ such that $f(r) = (r + I, r + J)$ induces an isomorphism*

$$R/IJ \cong (R/I) \times (R/J).$$

Before we begin the proof, a notational clarification must be made. We will borrow the *mod* notation from number theory. Let x and y be elements in the ring R. We write

$$x \equiv y \pmod{I},$$

to mean that $x - y$ lies in the ideal I. In other words, if Φ is the canonical homomorphism from R to R/I, then $\Phi(x) = \Phi(y)$.

Proof. First, we show the equality $IJ = I \cap J$. (Recall that in general, we have only that $IJ \subset I \cap J$.) Let $z \in I \cap J$. Since I, J are coprime, there exist $x \in I, y \in J$ such that $x + y = 1$. We have $z = z1 = zx + zy$. But since $z \in I \cap J$, it is also true that $zx \in JI = IJ$, and $zy \in IJ$. Thus $z \in IJ$.

Since f is a homomorphism, by basic isomorphism theorems, we have that

$$R/\ker f \cong f(R),$$

so it will suffice to show that f is surjective and that $\ker f = IJ$. The latter is straightforward: by definition of f, we have

$$\ker f = \{z \in R : z \in I, z \in J\} = I \cap J = IJ.$$

Let x, y be as above. Since $x + y = 1, y \in J$, we have $x = 1 \pmod J$. Similarly, $y \equiv 1 \pmod I$. So given any $r, s \in R$,

$$\begin{aligned} f(rx + sy) &\equiv (sy + I, rx + J) \\ &\equiv (s + I, r + J). \end{aligned}$$

Hence f is surjective, as desired.

We have an immediate generalization to finite collections of pairwise coprime ideals.

Definition. *Let R be a commutative ring. A collection $I_1, I_2, \ldots, I_t$ of ideals is* **pairwise coprime** *if for all $i \neq j$, $1 \leq i, j \leq t$, we have $I_i + I_j = R$.*

For such a collection of ideals, we have

$$R/(I_1 I_2 \cdots I_t) \cong (R/I_1) \times (R/I_2) \times \cdots (R/I_t).$$

The proof is a straightforward induction, but does require the following fact.

Proposition. *Given a collection $I_1, I_2, \ldots, I_{t-1}$ of ideals each of which are coprime to I_t, the product $J = I_1 I_2 \cdots I_{t-1}$ is coprime to I_t.*

Proof. To show that J and I_t are coprime, it suffices to show that the image of J in R/I_t by the natural projection map contains 1. By assumption, since each I_i is coprime to I_t, we can find in each I_i an element x_i whose image is 1. The product $x_1 x_2 \cdots x_{t-1}$ lies in the product J and its image in R/I_t is 1. Hence J and I_t are coprime.

A direct consequence of this proposition is that if I and J are coprime, then so are I^n and J^n, where

$$I^n = \overbrace{I \cdot I \cdots I}^{n \text{ times}}.$$

A few words of caution: there is the strong potential to misinterpret what I^2 ought to be, given an ideal I. It is *not* the ideal generated by the squares of elements in I! Also, if one views a ring R as an ideal of itself, then R^n as an ideal is not the same thing as R^n as an R-module. The notation here is ambiguous, yet unavoidably so, and therefore the reader should try to take care to make explicit what the exponent means if it isn't absolutely unmistakable.

We can use this result (along with the Chinese Remainder Theorem) to classify certain quotient rings.

Examples.

(i) Let n be a positive number in $\mathbf{Z}$, with its factorization,

$$n = \prod_{p,\,\text{prime}} p^{\alpha(p)}.$$

Then

$$\mathbf{Z}/n\mathbf{Z} \cong \prod_{p,\,\text{prime}} \mathbf{Z}/(p^{\alpha(p)}\mathbf{Z}).$$

(ii) Let f be a monic polynomial in the polynomial ring $K[x]$, with its factorization into irreducibles,

$$f = \prod_{g,\,\text{irred.}} g^{\alpha(g)}.$$

Then

$$K[x]/fK[x] \cong \prod_{g,\,\text{irred.}} K[x]/(g^{\alpha(g)}K[x]).$$

In the first example, this gives us a classification for the finite cyclic rings, and in the second this gives us a classification of the cyclic finite-dimensional K-algebras.

These rings, $\mathbf{Z}/(p^{\alpha_p}\mathbf{Z})$ and $K[x]/(g^{\alpha_g}K[x])$, are examples of *local rings.*

Definition. *Let R be a commutative ring. We call R a* **local ring** *if it has exactly one maximal ideal.*

Remark. An equivalent property is that $R \setminus R^*$ is an ideal of R. We leave the proof of the equivalence to the reader.

In fact, the rings in the above examples satisfy the much stronger condition that they have exactly one prime ideal. Therefore in these rings every element is either a unit or nilpotent, making these rings easy to work with, which is the motivation for the above decomposition.

8. Exercises

8.1. Let R and S be rings. An *R-S-bimodule* is an abelian group M that is provided both with a left R-module structure and with a right S-module structure, in such a way that $r(ms) = (rm)s$ for all $r \in R$, $m \in M$, $s \in S$. Prove that the category of R-S-bimodules is equivalent to the category of left modules over the ring $R \otimes S^{\mathrm{opp}}$; here S^{opp} is the ring S with the multiplication reversed, and the ring $R \otimes S^{\mathrm{opp}}$ is as defined in Exercise 6.15(a).

8.2. For a commutative ring R we denote, as in Exercise 1.33, by $B(R)$ the set of idempotents of R. From Exercises 1.33(b) and 1.34 we know that the set $B(R)$ can in a natural way be made into a Boolean ring.

(a) Prove that B extends to a covariant functor from the category **Crg** of commutative rings to the category **Bo** of Boolean rings, and that B has a left adjoint $H\colon \mathbf{Bo} \to \mathbf{Crg}$. What is $H(\mathbf{F}_2)$?

(b) Does B have a right adjoint? Does the inclusion functor $\mathbf{Bo} \subset \mathbf{Crg}$ have a right adjoint? a left adjoint?

8.3. Let k be a field, and let R be a *finite k-algebra*, i. e., a k-algebra that, when viewed as a vector space over k, has finite dimension. Let T be a sub-k-algebra of R. Prove that $T^* = T \cap R^*$.

8.4. Let R be a ring, and let $(M_i)_{i\in I}$ be a collection of R-modules. Prove that the direct sum $\bigoplus_{i\in I} M_i$ is the sum (in the categorical sense) of the modules M_i in the category ${}_R\mathbf{Mod}$ of R-modules, and that the product $\prod_{i\in I} M_i$ is the product (in the categorical sense) of the M_i in ${}_R\mathbf{Mod}$.

8.5. A module M over a ring is called *simple* if M has exactly two submodules, namely 0 and M itself.

(a) Let k be a field, n a positive integer, and let R be the ring $M(n,k)$ of $n \times n$-matrices over k. Exhibit a natural R-module structure on k^n, and show that k^n is simple as an R-module.

(b) Let R be a ring. In class it was shown that an R-module M is simple if and only if it is isomorphic to R/I for some maximal left ideal I of R. Give an example to show that I is not necessarily uniquely determined by M.

8.6. (a) Let R be a ring, let I be a left ideal of R, and let M be the R-module R/I. Prove that I is a two-sided ideal if and only if it is equal to the kernel of the ring homomorphism $R \to \operatorname{End} M$ defining the R-module structure on M.

(b) Suppose that M is a simple module over a *commutative* ring R. Prove that there is a *unique* maximal ideal I of R with $M \cong_R R/I$.

8.7. Let R be a ring, and let M be R, viewed as a left module over itself. Prove that the ring $\operatorname{End}_R M = \{f\colon M \to M : f \text{ is } R\text{-linear}\}$ is a subring of the ring of all endomorphisms of the additive group of M, and that $\operatorname{End}_R M$ is isomorphic to the ring R^{opp} opposite to R.

8.8. Let R be a ring, let I be a two-sided R-ideal, and let M be the R-module R/I. Prove that the group of R-automorphisms of M is isomorphic to the unit group of the ring R/I.

8.9. Let R be a ring. An R-module M is called *cyclic* if there exists $x \in M$ such that for every $y \in M$ there is an element $r \in R$ with $y = rx$.

(a) Prove: an R-module M is cyclic if and only if it is R-isomorphic to an R-module of the form R/I, where I is a left ideal of R.

(b) Let M be a cyclic R-module, and let $N \subset M$ be a submodule. Does it follow that N is cyclic? Does it follow that M/N is cyclic? Give a proof or a counterexample in each case.

8.10. Let R be a ring. An R-module M is said to be *finitely generated* if, for some non-negative integer n, there are $x_1, \ldots, x_n \in M$ such that for every $x \in M$ there are $r_1, \ldots, r_n \in R$ with $x = r_1x_1 + r_2x_2 + \cdots + r_nx_n$.

(a) Prove: a module over a field is finitely generated if and only if it has finite dimension as a vector space over that field.

(b) Prove: an R-module M is finitely generated if and only if for some non-negative integer n there is an exact sequence $R^n \to M \to 0$ of R-modules; here $R^n = \bigoplus_{i=1}^n R$.

8.11. Let R be a ring. An R-module M is called *noetherian* if every submodule of M is finitely generated.

(a) Give an example of a ring R and a finitely generated R-module that is not noetherian.

(b) Let $0 \to M_0 \to M_1 \to M_2 \to 0$ be a short exact sequence of R-modules. Prove: M_1 is noetherian if and only if both M_0 and M_2 are noetherian.

8.12. A ring is called *left noetherian* if it is noetherian when viewed as a left module over itself (see Exercise 8.11). Let R be a ring. Prove that the following four properties are equivalent:

(i) R is left noetherian;
(ii) every finitely generated R-module is noetherian;
(iii) for every chain $I_1 \subset I_2 \subset I_3 \subset \ldots$ of left ideals of R there exists an integer $m \geq 0$ such that for every $n \geq m$ one has $I_n = I_m$ (this is called the *ascending chain condition*);
(iv) for every non-empty set S of left ideals of R there exists $I \in S$ such that the only $J \in S$ with $I \subset J$ is given by $J = I$.

8.13. A commutative ring is called *noetherian* if it is left noetherian (see Exercise 8.12).

(a) Is $\mathbf{Z}$ noetherian? Is $\mathbf{Q}[X]$ noetherian?

(b) Give an example of a commutative ring that is not noetherian.

8.14. Let R be a ring. An R-module M is said to have *finite length* if, for some non-negative integer n, there is a chain

$$0 = M_0 \subset M_1 \subset \ldots \subset M_n = M$$

of submodules of M such that every module M_i/M_{i-1} $(0 < i \leq n)$ is simple. In that case n is called the *length* of M, notation: $\operatorname{length} M$ or $\operatorname{length}_R M$ or $l_R M$.

(a) Prove that $\operatorname{length} M$ is well-defined if it is finite, i. e., that it is independent of the choice of the chain of submodules.

(b) Prove that every module of finite length is finitely generated.

(c) Exhibit a finitely generated $\mathbf{Z}$-module that is not of finite length.

(d) Let M be an R-module, and let $N \subset M$ be a submodule. Prove that M has finite length if and only if both N and M/N have finite lengths, and that if these conditions are satisfied one has $\operatorname{length} M = \operatorname{length} N + \operatorname{length}(M/N)$.

8.15. (a) Let R be a non-zero ring. Prove that there is a simple R-module.

(b) A *skew field* (or *division ring*) is a ring k whose unit group is equal to $k - \{0\}$. Describe all simple R-modules when R is a skew field, when $R = \mathbf{Z}$, and when $R = k[X]$ for a field k.

8.16. (a) Let R be a ring. Prove that every finitely generated R-module has finite length if and only if R has finite length when viewed as a left module over itself.

(b) Give an example of a ring satisfying the condition in (a) that is not a field.

(c) Let V be a finite dimensional vector space over a field k. Prove that $\operatorname{length}_k V$ exists and is equal to $\dim_k V$.

8.17. (a) Let A be an abelian group. Prove that A has finite length as a $\mathbf{Z}$-module if and only if A is finite, and express $\operatorname{length}_{\mathbf{Z}} A$ in terms of $\#A$ if A is finite.

(b) Let $k[X]$ be the polynomial ring in one indeterminate X over a field k, and let M be a $k[X]$-module. Prove: M has finite length over $k[X]$ if and only if M has finite dimension when viewed as a k-vector space. Prove also that in that case one has $\operatorname{length}_{k[X]} M \leq \dim_k M$; for which fields is it true that one always has equality here?

8.18. (a) A short exact sequence $0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$ of modules over a ring R is said to *split* if there is an R-homomorphism $h: A \oplus C \to B$ such that $hi = f$ and $gh = \pi$, where $i: A \to A \oplus C$ is the canonical injection and $\pi: A \oplus C \to C$ is the canonical projection. Prove that any such h is automatically an isomorphism.

(b) Let $0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$ be a short exact sequence of modules over a ring R. Prove that the sequence splits if and only if there is an R-homomorphism $i: C \to B$ with $gi = 1_C$, and if and only if there is an R-homomorphism $p: B \to A$ with $pf = 1_A$.

8.19. (a) Prove that every short exact sequence of modules over a skew field splits (see Exercises 8.15 and 8.18 for the definitions).

(b) Give an example of a ring R and a short exact sequence of R-modules that does not split.

(c) Let $0 \to A \to B \to C \to 0$ be a short exact sequence over a ring R with $C \cong_R R$. Prove that the sequence splits.

8.20. Let R_1 and R_2 be rings, and let R be the product ring $R_1 \times R_2$.

(a) Prove: if M_i is an R_i-module, for $i = 1, 2$, then $M = M_1 \times M_2$ is in a natural manner an R-module, and every R-module is isomorphic to one of this form.

(b) Let M_i and N_i be R_i-modules, for $i = 1, 2$, and let M and N be the R-modules $M_1 \times M_2$ and $N_1 \times N_2$, respectively. Prove that the natural map ${}_{R_1}\mathrm{Hom}(M_1, N_1) \times {}_{R_2}\mathrm{Hom}(M_2, N_2) \to {}_R\mathrm{Hom}(M, N)$ is a bijection.

Note. In categorical language one may express the results of this exercise by saying that the assignment $(M_1, M_2) \mapsto M_1 \times M_2$ provides an equivalence between a suitably defined "product category" ${}_{R_1}\mathbf{Mod} \times {}_{R_2}\mathbf{Mod}$ and ${}_R\mathbf{Mod}$, the notation being as in Exercise 8.4.

8.21. Let R be a ring. If M is an R-module and $x \in M$, then the *annihilator* of x in R is defined by $\mathrm{Ann}_R\, x = \{r \in R : rx = 0\}$. It is the kernel of the R-homomorphism $R \to M$ mapping 1 to x, so it is a left ideal of R. Let now M be a *simple* R-module.

(a) Prove: $\mathrm{Ann}_R\, x$ is a *maximal* left ideal of R for every $x \in M - \{0\}$, and $I = \bigcap_{x \in M} \mathrm{Ann}_R\, x$ is a two-sided ideal of R. Prove also that R/I is isomorphic to a subring of the ring $\mathrm{End}\, M$ of all endomorphisms of the abelian group M.

(b) Prove that the ring $\mathrm{End}_R\, M$ of R-linear endomorphisms of R is a skew field (as defined in Exercise 8.15(a)).

8.22. Let R be a ring.

(a) Prove that the intersection of all maximal left ideals of R is a two-sided ideal of R.

(b) Let I be a maximal left ideal of R, and put $T = \{r \in R : Ir \subset I\}$. Prove that T is a subring of R containing I, that I is a two-sided ideal of T, and that T/I is a skew field.

8.23. Let R be a non-zero ring. Prove: R is a skew field if and only if every R-module is free, and if and only if every simple R-module is free.

9. Finite Algebras

In the previous section, we noted a certain similarity in the classification of certain finite cyclic rings and of certain cyclic finite-dimensional K-algebras. In both cases, the condition requiring finiteness turns out to be crucial. We now prove a theorem which is a generalization of the Chinese Remainer Theorem which holds for these special cases of algebras. This theorem will be useful in the later context of Galois theory when we consider finite field extensions, which are special cases of finite K-algebras.

Definition. *Let K be a field. A* **finite K-algebra** *is a K-algebra A which is finite-dimensional as a vector space over K.*

Theorem. *Let A be a commutative ring that is either finite or a finite K-algebra for some field K. Then $\mathrm{Spec}(A)$ is finite (as a set), each $M \in \mathrm{Spec}(A)$ is maximal, and there exist positive integers $n(M)$, for $M \in \mathrm{Spec}(A)$, such that*

$$A \cong \prod_{\mathrm{Spec}(A)} (A/M^{n(M)}),$$

as an isomorphism of rings by the natural coordinate-wise projection map. Moreover, each $A/M^{n(M)}$ is a local ring, with maximal ideal $M/M^{n(M)}$, and each element of $A/M^{n(M)}$ is either a unit or nilpotent.

Proof. We begin with a preliminary claim that if such an A as above is a domain, then it must be a field. To see this, consider a non-zero element $x \in A$ and the right multiplication map $y \mapsto yx$. Since A is a domain, x is not a zero-divisor and this map is injective. If A is finite (as a set), then this map must also be surjective, so there exists y such that $yx = 1$ and hence x is a unit. If A is a finite-dimensional K-algebra, in particular it is a vector space over K. The map is injective since A is a domain, as before. Since the multiplication map is K-linear, the image $xA \subset A$ is a subspace of A as a K-vector space. If the map is not surjective, then by dimension-counting, we have a nontrivial kernel, but the map is injective. Hence the map is surjective, and there exists $y \in A$ such that $xy = 1$, and x is a unit, as required. Hence A is a field.

We now show that every $M \in \mathrm{Spec}(A)$ is in fact maximal. Given a prime ideal $M \in \mathrm{Spec}(A)$, we have that A/M is a domain. Moreover, since A is finite or a finite-dimensional K-algebra and the projection maps are surjective, A/M is also finite or a finite-dimensional K-algebra. So by the previous claim, A/M is a field, and hence M must be maximal.

We show that $\mathrm{Spec}(A)$ is finite by exhibiting an upper bound on the number of (pairwise distinct) maximal ideals $M_1, M_2, \ldots, M_t$ in A. Recall that any two distinct maximal ideals are coprime. Hence by the Chinese Remainder Theorem, we have a surjective map

$$A \to \prod_{M} (A/M).$$

Recall that each A/M_i is a field. Consider the case where A is finite, as a set. Since fields are non-zero, in particular, they must contain at least 2 elements. We have the inequality

$$|A| \geq 2^t$$

given t distinct maximal ideals. Since A is finite, we have an upper bound on t, namely, $t \leq \log_2 |A|$. Consider the case where A is a finite-dimensional K-vector space. Then each A/M is also a finite-dimensional K-vector space, and since it is non-zero, is of dimension at least 1. Thus we have $\dim(A) \geq t$, again an upper bound on t. Hence $\operatorname{Spec}(A)$ is finite, in either case.

We now study the kernel of the (natural coordinate-wise projection) map

$$A \to \prod_M (A/M),$$

where the product is over all maximal ideals M in A. The kernel is clearly the intersection $\bigcap_M M$ of all maximal ideals. However, we have already shown that for the given A, any prime ideal is maximal (and vice versa). Recall that the intersection of all prime ideals is the nilradical of A, the set of all nilpotent elements of A.

Claim. *The nilradical, $\sqrt{0_A}$, is itself nilpotent as an ideal.*

Proof. There are two cases. If A is finite, the nilradical is also finite. If A is a finite K-algebra, since any ideal of A is a submodule over K, in particular it is a subspace of a finite-dimensional vector space, and hence finite-dimensional. In either case, there exists a finite set $b_1, \dots, b_s$ of elements in the nilradical such that

$$\sqrt{0_A} = Ab_1 + Ab_2 + \cdots + Ab_s.$$

We say that $\sqrt{0_A} = (b_1, \dots, b_s)$, or that it is the ideal *generated* by $\{b_1, \dots, b_s\}$.

It is easy to see that a power of the nil radical, $\sqrt{0_A}^N$ is generated by elements of the form

$$\prod_{i=1}^{s} b_i^{n_i} \quad \text{such that} \quad \sum_{i=1}^{s} n_i = N.$$

Since each of the b_i is nilpotent, for each there is an l_i such that $b_i^{l_i} = 0$. We can take N to be

$$N = 1 + \sum_{i=1}^{s} (l_i - 1).$$

We can now take any generator of $\sqrt{0_A}^N$ in the above form, $\prod b_i^{n_i}$. Since the n_i must add up to N, by the pigeonhole principle, there must be some i for which $n_i \geq l_i$. For this

i, $b_i^{n_i} = 0$, and so the entire product is 0. This shows that every generator of $\sqrt{0_A}^N$ is 0, and so $\sqrt{0_A}^N = \{0\}$. This proves the claim.

This being so, we notice that

$$\{0\} = \sqrt{0_A}^N = (\bigcap_M M)^N = (\prod_M M)^N = \prod_M (M^N)$$

by making use of the fact that maximal ideals are necessarily coprime, and the last equality holds from associativity of ideal multiplication. Recall that if two ideals are coprime, so are their powers (a corollary of an earlier proposition). Hence all the M^N, M maximal, are coprime. By the Chinese Remainder Theorem, we have that

$$A \cong A/\{0\} \cong A/(\prod_M M^N) \cong \prod_M (A/M^N).$$

Further, we know that for each M,

$$(A/M) = (A/M^N)/(M/M^N),$$

and since A/M is a field, M/M^N must maximal in A/M^N.

By a previous proposition on maximal ideals, we know that

$$\mathrm{Spec}(A) = \coprod_M \mathrm{Spec}(A/M^N).$$

We count the number of elements in $\mathrm{Spec}(A/M^N)$ for each maximal ideal M. The number of ideals in $\mathrm{Spec}(A)$ is the number of M, and since the above union is disjoint, each set $\mathrm{Spec}(A/M^N)$ must therefore have exactly one element. So each A/M^N has exactly one prime (maximal) ideal, namely M/M^N. In addition, we observe that every element is either a unit (if it is outside this prime ideal) or nilpotent (if it is inside).

One might suspect that this talk of finite sets and finite algebras may have a further generalization of some sort. Indeed, this is the case, and the above theorem holds for any commutative *Artinian ring*, that is a ring which satisfies the "Descending Chain Condition" (DCC) for ideals: every descending chain of ideals in the ring terminates.

9. Exercises

9.1. Let p be a prime number. Prove that every ring of cardinality p^2 is commutative. How many are there, up to isomorphism? Prove the correctness of your answer.

9.2. Construct a commutative ring R with the property that for all positive integers n one has $\sqrt{0_R}^n \neq (0)$.

9.3. Let R be a commutative ring. We call R an *Artin ring*, or *artinian*, if for every chain

$$I_1 \supset I_2 \supset I_3 \supset \ldots$$

of ideals of R there exists an integer $m \geq 0$ such that for every $n \geq m$ one has $I_n = I_m$ (this is called the *descending chain condition*).

(a) Prove: R is an Artin ring if and only if for every non-empty set S of ideals of R there exists $I \in S$ such that the only $J \in S$ with $J \subset I$ is given by $J = I$.

(b) Prove: a domain is artinian if and only if it is a field.

(c) Is $\mathbf{Z}$ artinian? Is $\mathbf{Q}[X]/(X^{12} - 729)^2\mathbf{Q}[X]$ artinian?

(d) Prove: every prime ideal of an Artin ring is maximal.

9.4. Let R be a commutative Artin ring (see Exercise 9.3).

(a) Prove that R has only finitely many maximal ideals.

(b) Let $J(R)$, as in Exercise 1.29, be the intersection of all maximal ideals of R. Prove that $J(R)$ is nilpotent, i.e., that $J(R)^n = 0$ for some integer $n \geq 0$. (*Hint*: first find n such that $I = J(R)^n$ satisfies $I^2 = I$, and derive a contradiction from $I \neq 0$ by considering a minimal principal ideal Ra with $aI \neq 0$.)

(c) Prove: there is a finite sequence of *local* Artin rings $R_1, \ldots, R_t$ with nilpotent maximal ideals such that R is isomorphic to $\prod_{i=1}^{t} R_i$.

9.4a. Let R be a local ring, $\mathfrak{m}$ its maximal ideal, and $k = R/\mathfrak{m}$ its residue class field. For a non-negative integer n, define $\mathfrak{m}^n$ by $\mathfrak{m}^0 = R$ and $\mathfrak{m}^{i+1} = \mathfrak{m} \cdot \mathfrak{m}^i$.

(a) Let n be a non-negative integer. Prove that the R-module $\mathfrak{m}^n/\mathfrak{m}^{n+1}$ has the structure of a vector space over k.

(b) Suppose in addition that R is *finite*. Prove that the cardinality of R is a power of the cardinality of k.

9.5. Let R be a commutative ring. Prove: R is an Artin ring (as defined in Exercise 9.3) if and only if R has finite length as a module over itself. Prove also that every commutative Artin ring is noetherian (as defined in Exercise 8.13).

9.6. Let k be a field, let C be a group of order 3, and let A be the group ring $k[C]$.

(a) Prove that A is a finite commutative k-algebra.

(b) By (a) and a theorem proved in class, A can be written as the product of t local k-algebras with nilpotent maximal ideals, for some non-negative integer t. Prove that one has $t \in \{1, 2, 3\}$. Also, describe for which fields k one has $t = 1$, for which fields one has $t = 2$, and for which fields one has $t = 3$. Do all three values of t actually occur?

(c) For which fields k is A a product of *fields*?

9.7. An algebra A over a ring R is said to be *totally split* if for some non-negative integer n there is an R-algebra isomorphism $A \to \prod_{i=1}^{n} R$, where $\prod_{i=1}^{n} R$ has componentwise ring operations.

Let k be a finite field, and suppose that $f \in k[X]$ is a non-zero polynomial for which the k-algebra $k[X]/fk[X]$ is totally split. Prove: $\deg f \leq \#k$.

9.8. Let G be a finite abelian group, let $\mathbf{C}$ be the field of complex numbers, and let $\mathbf{C}[G]$ be the group ring of G over $\mathbf{C}$.

(a) Exhibit a bijection between the set of $\mathbf{C}$-algebra homomorphisms $\mathbf{C}[G] \to \mathbf{C}$ and prove the set of group homomorphisms $G \to \mathbf{C}^*$, and that the cardinality of these sets equals $\#G$.

(b) Prove that $\mathbf{C}[G]$ is totally split as a $\mathbf{C}$-algebra (see Exercise 9.7).

9.9. Let G be a finite abelian group, let $\mathbf{R}$ be the field of real numbers, and let $\mathbf{R}[G]$ be the group ring of G over $\mathbf{R}$. Prove that there are integers $n > 0$ and $m \geq 0$ with $n + 2m = \#G$ such that there is an $\mathbf{R}$-algebra isomorphism $\mathbf{R}[G] \cong \mathbf{R}^n \times \mathbf{C}^m$. Can you find a formula for n?

9.10. Let **Fb** be the category of finite Boolean rings, and let **sets** be the category of finite sets. Prove that the functors **Fb** $\to$ **sets** and **sets** $\to$ **Fb** sending R to $\operatorname{Spec} R$ and X to $P(X)$ (see Exercise 1.36), respectively, are anti-equivalences of categories.

9.11. Let p be a prime number, k a field of characteristic p, and G a finite abelian p-group. Prove that the group ring $k[G]$ is local.

10. Homomorphism modules and tensor products

We recall that in Section 8 we defined the notion of a free module. Given a set S, the free R-module, $F(S) = R^{(S)}$, is

$$\bigoplus_{i \in S} R.$$

One can ask in general if a given module, M, can be considered "free." In essence, the issue is whether M is isomorphic to a module of the above form. The answer to whether a module is free lies in the existence (or lack thereof) of a basis.

Definition. *Given an R-module, M, we say that a subset $S \subset M$ is a* **basis** *of M if every element x of M can be written uniquely as*

$$x = \sum_{s \in S} r_s s$$

where each r_s is in R and almost all r_s are equal to zero (so that the sum makes sense).

There are two parts to this definition. If every element $x \in M$ can be written in this form, we say that S *spans* M. If the way of writing any given $x \in M$ in this form is unique, we say that the elements of S are *linearly independent*. So in other words, a basis is a linearly independent subset of M which spans M.

Definition. *An R-module is called* **free** *if it has a basis.*

For example, let K be a field. A K-module is a vector space over K, which we know from linear algebra must have a basis (even if it is not finite-dimensional). Therefore, all K-modules are free.

If M has a basis S, then $M \cong_R R^{(S)}$, via the isomorphism

$$\sum_{s \in S} r_s s \longleftrightarrow \langle r_s \rangle_{s \in S}$$

which the reader can verify to be an isomorphism.

The module R^n is called the *free module of rank n*. This leads us to the natural question: If $R^n \cong_R R^m$, then can we conclude that $n = m$? The answer in general is no, for with non-commutative rings counterexamples can be found. We know from linear algebra that if R is a field, the answer is yes. What about for other commutative rings? If $R = \{0\}$, then the answer is quite clearly no, but for all other commutative rings, the problem may be reduced to the case for fields by factoring through a maximal ideal.

In fact, we can ask a more general question about free modules. Namely, if we have two sets, S and T, such that

$$F(S) \cong_R F(T), \quad \text{i.e.} \quad R^{(S)} \cong_R R^{(T)},$$

can we conclude that $\#S = \#T$? Again, if R is a field, we know from linear algebra that this is true. What about general rings? As seen above, if $\#S$ is finite, then this depends on whether the ring is commutative. However, if $\#S$ is infinite, then regardless of whether the ring is commutative, the answer is yes. We have the following lemma.

Lemma. *Let R be a non-zero ring, and let S and T be sets such that at least one of them is infinite. Let $R^{(S)} \cong_R R^{(T)}$. Then $\#S = \#T$.*

Proof. Let Φ be the isomorphism from $R^{(S)}$ to $R^{(T)}$. We examine what this does to the basis of $R^{(S)}$. Each basis element s is mapped to a linear combination of a finite subset T_s

of T. Yet the union of these sets T_s spans $R^{(T)}$, so $T = \bigcup_{s\in S} T_s$. Hence if S is finite, then T is also finite. By symmetry, if either of these sets (T or S) is finite, then so is the other, and therefore if either of these sets is infinite, then so is the other. Our condition was that at least one of these is infinite, so we can conclude that they both are. Furthermore, from $T = \bigcup_{s\in S} T_s$ it follows that

$$\#T = \#\big(\bigcup_{s\in S} T_s\big) \leq \#S \cdot \#\mathbf{N} = \#S,$$

and by symmetry, $\#S \leq \#T$. Therefore, $\#S = \#T$.

In what follows, we let R be an arbitrary ring, not necessarily commutative. Let M, N be left R-modules. Then the set of R-linear maps $\mathrm{Hom}_R(M, N)$ has the structure of an abelian group, with addition defined pointwise, $(f + g)(x) = f(x) + g(x)$. (Since $r(f + g)(x) = rf(x) + rg(x) = f(rx) + g(rx) = (f + g)(rx)$, this map is still R-linear.)

In the case where R is commutative, this abelian group has the additional structure of an R-module. There are two different ways of defining the action of R on an element $f \in \mathrm{Hom}_R(M, N)$, but they turn out to be the same. Namely, we have

$$\begin{array}{ccc} R \times \mathrm{Hom}_R(M, N) & \to & \mathrm{Hom}_R(M, N) \\ (r, f) & \mapsto & \left(rf\colon x \mapsto \begin{cases} \text{(i)} & r(f(x)) \\ \text{(ii)} & f(rx) \end{cases} \right), \end{array}$$

where the first definition $rf\colon x \mapsto r(f(x))$ uses the R-module structure of N, and the latter uses that of M. These are equivalent because the map f is R-linear. We still need to check, however, that this rf is R-linear, and to show this we need to use the commutativity of R. Namely,

$$rf(sx) = r(sf(x)) = (rs)f(x) = (sr)f(x) = s(rf(x)).$$

In the case where R is non-commutative, the group of homomorphisms $\mathrm{Hom}_R(M, N)$ acquires a module structure if we assume that the modules M, N have some additional structure. First, we need a definition.

Definition. *An abelian group M is an* R-S-**bimodule**, *where R, S are arbitrary rings, if it is*

(i) a left R-module;

(ii) a right S-module (with the **same** *additive group);*

(iii) it satisfies the compatibility condition $(rm)s = r(ms)$, for all $r \in R$, $m \in M$, $s \in S$.

Notice that any R-S-bimodule is automatically a left R-module and a right S-module. Also, any left R-module is a R-$\mathbf{Z}$-bimodule, and similarly any right S-module is a $\mathbf{Z}$-S-bimodule. (These are the default rings; remember that any abelian group is naturally a

Z-module.) The simplest example of a bimodule is the ring R itself: any ring R is an R-R-bimodule.

Now suppose that M is an R-S-bimodule, and N is an R-T-bimodule. We can give $\mathrm{Hom}_R(M, N)$ the structure of an S-T-bimodule in the following natural way:

$$(ft)(x) = f(x)t, \qquad (sf)(x) = f(xs).$$

It is left as an exercise for the reader to verify that these definitions do indeed yield a bimodule structure.

If R is commutative, then every left R-module can be made a right R-module by defining $mr = rm, r \in R, m \in M$. It is left to the reader to verify that this is indeed a right R-module structure on the underlying abelian group of M. (Note, however, that the commutativity of R *is* used in the verification!) This right module structure is compatible with the original left R-module structure, so in fact, M can be regarded as an R-R-bimodule. The compatibility condition simply reduces to the statement that $s(rm) = (sr)m = (rs)m = r(sm)$. (However, for most commutative rings R, not every R-R-bimodule arises in this way.)

The "Hom" functor is also *left exact*, namely, if

$$0 \to N_1 \to N_2 \to N_3$$

is exact, then

$$0 \to \mathrm{Hom}_R(M, N_1) \to \mathrm{Hom}_R(M, N_2) \to \mathrm{Hom}_R(M, N_3)$$

is also an exact sequence of abelian groups, of R-modules (if R is commutative), or of S-T-bimodules (if M is an R-S-bimodule, the N_i are R-T-modules), depending on the context. Similarly, if

$$M_1 \to M_2 \to M_3 \to 0$$

is exact, then

$$0 \to \mathrm{Hom}_R(M_3, N) \to \mathrm{Hom}_R(M_2, N) \to \mathrm{Hom}_R(M_1, N)$$

is also exact.

Since $\mathrm{Hom}_R(R, N)$ is an R-$\mathbf{Z}$-bimodule, it is in particular a left R-module. In this special case where $M = R$, and we view $\mathrm{Hom}_R(R, -)$ as a functor on left R-modules, we see that we have an isomorphim $\mathrm{Hom}_R(M, N) \cong N$, by the following map: $f \mapsto f(1)$. So the functor $\mathrm{Hom}_R(R, -)$ is exact, not just left or right exact. (One needs to check that the isomorphism given above is compatible with maps between the N_i.) This is actually true for any free R-module $M = R^{(S)}$. Namely, we have the isomorphism $\mathrm{Hom}_R(R^{(S)}, N) = N^S$, and one can check that exact sequences are preserved.

Definition. *Let M be an R-module. Then M is* **projective** *if the functor ${}_R\mathrm{Hom}(M, -)$ is exact.*

As we have just seen, every free module is projective. Further investigation reveals that while not every projective module is free, that for every projective module M there is a module N such that $M \oplus N$ is free. (A book entitled *Serre's Conjecture* (which has now become a theorem) by T.-Y. Lam discusses the conditions under which projective modules are and are not free in greater depth.) There is a parallel notion for the contravariant case.

Definition. *Let N be an R-module. Then N in* **injective** *if the functor ${}_R\mathrm{Hom}(-, N)$ is exact.*

While projective modules are quite easy to construct with free modules, injective modules in general require more effort. An example of an injective module is provided here for flavor:

Example. An abelian group G is called *divisible* if the map $n\colon G \to G$ (which takes $x \mapsto nx$) is surjective. A $\mathbf{Z}$-module is injective if and only if it is divisible as an abelian group.

We now shift our attention to tensor products. If R is any ring, and M is a right R-module, and N is a left R-module, then we define the tensor product

$$M \otimes_R N$$

to be the abelian group generated by elements of the form $x \otimes y$ with $x \in M$ and $y \in N$, and with relations

(i) $(x_1 + x_2) \otimes y = (x_1 \otimes y) + (x_2 \otimes y)$;

(ii) $x \otimes (y_1 + y_2) = (x \otimes y_1) + (x \otimes y_2)$;

(iii) $(xr) \otimes y = x \otimes (ry)$;

for all $r \in R$, $x, x_1, x_2 \in M$, and $y, y_1, y_2 \in N$.

Definition. *Let M and N be right and left R-modules respectively, and let C be an abelian group. Let $f\colon M \times N \to C$ be a map. We say that f is R*-**bilinear** *if for all m, m_1, $m_2 \in M$, n, n_1, $n_2 \in N$, and $r \in R$, we have*

(i) $f(m_1 + m_2, n) = f(m_1, n) + f(m_2, n)$.

(ii) $f(m, n_1 + n_2) = f(m, n_1) + f(m, n_2)$.

(iii) $f(mr, n) = f(m, rn)$.

So the tensor map $\otimes : M \times N \to M \otimes_R N$ is R-bilinear, and if C is an abelian group, then

$$\mathrm{Hom}(M \otimes_R N, C) \cong \mathrm{Bil}_R(M \times N, C),$$

where Bil_R denotes the set of R-bilinear maps. If M is an S-R-bimodule and N is an R-T-bimodule, then $M \otimes_R N$ is an S-T-bimodule, by multiplying as follows:

$$\begin{array}{rcl} s(x \otimes y) & = & (sx) \otimes y \\ (x \otimes y)t & = & x \otimes (yt) \end{array}$$

When R is commutative, this simplifies our picture somewhat. If M and N are R-modules (and therefore R-R-bimodules), then $M \otimes_R N$ is an R-module. Another interesting thing to notice is that for any ring R (even non-commutative), we have

$$R \otimes_R M \cong M,$$

generalizing what we noticed earlier, taking the tensor products of abelian groups and finding that $\mathbf{Z} \otimes G \cong G$. We note, for emphasis, some important facts and examples of tensor products.

(i) Let M be a right R-module, and N a left R-module. The tensor product over R of the two modules, $M \otimes_R N$, is an abelian group.

(ii) Let R be a commutative ring and let M and N be R-modules. The tensor product $M \otimes_R N$ has the additional structure of an R-module.

In the remainder of this section, we collect some facts about tensor products which will become quite useful in the context of Galois theory. If not explicitly stated, an "R-module" is a "left R-module."

Fact (i). Let M be a right R-module, N be a left R-module. The functors $- \otimes_R N$ and $M \otimes_R -$ are *right* exact functors. That is, the functor $- \otimes_R N$ takes a right exact sequence of right R-modules to a right exact sequence of abelian groups, and the functor $M \otimes_R -$ takes a right exact sequence of left R-modules to a right exact sequence of abelian groups.

Fact (ii). We have a "distributive law" for direct sums and tensor products, namely

$$\Big(\bigoplus_{i \in I} M_i\Big) \otimes_R N \cong \bigoplus_{i \in I} (M_i \otimes_R N),$$

by the natural map. (The proof mainly consists of checking that this map is well-defined.)

Fact (iii). One can view the next property as a generalization of the first two properties. It will be stated using the language of categories. Namely, the functors $- \otimes_R N$ and $M \otimes_R -$ commute with arbitrary right limits. This can be proved by invoking one of the exercises, in which it is shown that a functor that has a right adjoint preserves right limits, and that a functor that has a left adjoint preserves left limits.

Fact (iv). The following property is crucial in that it allows us, on occasion, to get rid of a tensor product sign. Namely,

$$R \otimes_R N \cong N,$$

through the natural map $r \otimes x \mapsto rx$, and the inverse map is $x \mapsto 1 \otimes x$. The isomorphism is not just as abelian groups, but also as R-modules. Note, by the way, that in many of these properties we have actually a natural map which does the job of the isomorphism.

Fact (v). If I is a right ideal in R, recall that R/I is a right module over R. This property is a generalization of an observation made in the case of tensor products of abelian groups, i.e. $\mathbf{Z}/n\mathbf{Z} \otimes A \cong A/nA$, where A an abelian group. In the setting of modules, we have

$$(R/I) \otimes_R N \cong N/IN,$$

where we define the product of an ideal I of R with the R-module N to be the subgroup of N generated by the set $\{xy : x \in I, y \in N\}$.

Proof. We sketch the proof. We use the right-exactness of the functor $- \otimes_R N$. Namely, we have the right exact sequences

$$\begin{array}{ccccccc} I & \to & R & \to & R/I & \to & 0, \\ I \otimes_R N & \to & R \otimes_R N & \to & (R/I) \otimes_R N & \to & 0. \end{array}$$

We already know from the previous property that $R \otimes_R N \cong N$, and since the sequence is exact, we only need that the image of $I \otimes_R N$ is exactly IN. The map from $I \otimes_R N$ to $R \otimes_R N$ is the inclusion map taking $x \otimes_R y$ as an element of $I \otimes_R N$ to $x \otimes_R y$ as an element of $R \otimes_R N$. The canonical map from $R \otimes_R N$ to N takes $x \otimes y$ ($x \in I$, $y \in N$) to the element $xy \in IN$. Since IN is exactly generated by these elements, the image is IN and we are done.

Fact (vi). For the statement of this property, let S be any set (not a ring, as it often is). Then

$$R^{(S)} \otimes_R N \cong N^{(S)},$$

where $N^{(S)}$ denotes a direct sum of S copies of N. This follows from properties (iv) and (ii). Note that we use our explicit map from $R \otimes_R N \mapsto N$ to construct the isomorphism.

Fact (vii). For some R-modules, M, the functor $M \otimes_R -$ is not only right exact, but is left exact as well. An R-module is called *flat* if the functor $M \otimes_R -$ is exact.

In particular, all free modules are flat. This can be seen quite easily by noticing that if $M \cong R^{(S)}$, we have

$$M \otimes_R - \cong (-)^{(S)},$$

which is trivially an exact functor. Lang shows that a module is flat if and only if it is a direct limit of free modules. (Chapter XVI, Exercise 13).

We can see an example of a non-projective flat module even when we take the ring R to be $\mathbf{Z}$. An example of a flat $\mathbf{Z}$-module which can be shown to be non-projective is the additive group $\mathbf{Q}$ of rational numbers. However, any flat $\mathbf{Z}$-module that is finitely generated is projective.

Another quick interesting example comes if we let $R = \mathbf{R}$. Since $\mathbf{C}$ is a free module over $\mathbf{R}$, and therefore flat, it follows that $\mathbf{C} \otimes_{\mathbf{R}} -$ is an exact functor.

Fact (viii). Let M be an S-R-bimodule, N be an R-T-bimodule, and P be a T-U-bimodule. Then

$$(M \otimes_R N) \otimes_T P \quad {}_S\cong_U \quad M \otimes_R (N \otimes_T P).$$

Fact (ix). If R is commutative, then $M \otimes_R N \cong N \otimes_R M$.

Fact (x). Let A be an R-algebra, an N be an R-module. Then $A \otimes_R N$ has a natural A-module structure. The additive group is clear, and scalar multiplication goes by

$$a_1(a_2 \otimes n) = (a_1 a_2) \otimes n.$$

Fact (xi). Let A and B be R-algebras. The tensor product $A \otimes_R B$ is an R-algebra. We already have the additive structure. The ring structure can be defined through coordinate-wise multiplication:

$$(x \otimes y)(z \otimes w) = (xz) \otimes (yw),$$

for all $x, z \in A$ and $y, w \in B$. (Recall that one must use the universal property of tensor products to verify that this is a well-defined map on a general element of the tensor product.) The map $f' = f_{A \otimes_R B}$ from R into $A \otimes_R B$ is given by the diagram below:

$$\begin{tikzcd} A \arrow[r, "- \otimes 1"] & A \otimes_R B \\ R \arrow[u, "f_A"] \arrow[ur, "f'"] \arrow[r, "f_B"'] & B \arrow[u, "1 \otimes -"'] \end{tikzcd}$$

Next suppose that R, A, and B are commutative. Then for any commutative R-algebra C and any pair of R-algebra maps $\alpha\colon A \to C$ and $\beta\colon B \to C$ there is a unique R-algebra map $\Phi\colon (A \otimes_R B) \to C$ such that $\alpha = \Phi \circ (- \otimes 1)$ and $\beta = \Phi \circ (1 \otimes -)$. Hence if R is commutative, then $- \otimes_R -$ is a *coproduct* in the category of commutative R-algebras.

10. Exercises

10.1. Let R be a ring. An R-module is called *free* if for some set S it is isomorphic to $R^{(S)}$.

(a) Suppose that R is non-zero and that S is a set. Prove that $R^{(S)}$ is a finitely generated R-module if and only if S is finite.

(b) Let M be an R-module and let S be a set. Prove that there is a bijection ${}_R\mathrm{Hom}(R^{(S)}, M) \to M^S$ that is functorial in M. Are your bijections group isomorphisms?

(c) Let F be a free R-module. Prove that every short exact sequence $0 \to M \to N \to F \to 0$ of R-modules splits (as defined in Exercise 8.18).

10.2. (a) Let k be a field. Prove that every short exact sequence of k-modules splits.

(b) Let R be a finite commutative ring with $\sqrt{0_R} = 0$. Prove that every short exact sequence of R-modules splits.

(c) Let R be a commutative ring with $\sqrt{0_R} \neq 0$. Prove that there is a short exact sequence of R-modules that does not split.

10.3. For a ring R, we denote by R^n the direct sum of n copies of R; this is an finitely generated free R-module.

(a) Suppose that R is non-zero and commutative, and let n and m be non-negative integers. Prove: if $R^n \cong R^m$ as R-modules, then $n = m$. (*Hint*: reduce to the case that R is a field.)

(b) Construct a non-zero abelian group A for which $A \oplus A \cong A$ (as abelian groups), and prove that for any such A the ring $R = \mathrm{End}\, A$ has the property that $R \cong R^2$ as R-modules, and $R \neq \{0\}$.

10.4. Let R be a ring, and let P be an R-module. Recall that P is called *projective* if the functor $M \mapsto \mathrm{Hom}(P, M)$ from the category of R-modules to the category of abelian groups is exact.

(a) Prove that the following properties of P are equivalent:

(i) P is projective;

(ii) for every surjective map $f\colon M \to M''$ of R-modules, the map $f_*\colon {}_R\mathrm{Hom}(P, M) \to {}_R\mathrm{Hom}(P, M'')$ is surjective;

(iii) every short exact sequence $0 \to M' \to M \to P \to 0$ of R-modules splits;

(iv) there is an R-module Q for which $P \oplus Q$ is free.

(b) Suppose that P is finitely generated and projective. Prove that there is a finitely generated projective R-module Q such that $P \oplus Q \cong R^n$ for some non-negative integer n.

10.5. Let R be a ring. Recall that an R-module M is called *flat* if $- \otimes_R M$ is an exact functor from the category of right R-modules to the category of abelian groups. Prove that every projective R-module is flat.

10.6. (a) Prove that a $\mathbf{Z}$-module is projective if and only if it is free.

(b) Prove that a **Z**-module is flat if and only if it is torsionfree (i. e., has no non-zero elements of finite order).

10.7. Let R, S, T, U be rings, M an R-S-bimodule, N an S-T-bimodule, and P a U-T-bimodule. Exhibit an isomorphism

$$\mathrm{Hom}_T(M \otimes_S N, P) \cong \mathrm{Hom}_S\big(M, \mathrm{Hom}_T(N, P)\big)$$

of U-R-bimodules that is functorial in each of M, N, and P. You do not need to check that your map is well-defined, that it is an isomorphism, or that it is functorial; but do list *what* needs to be checked.

Note. Similarly, if Q is an R-U-bimodule, there is a functorial isomorphism

$${}_R\mathrm{Hom}(M \otimes_S N, Q) \cong {}_S\mathrm{Hom}\big(N, {}_R\mathrm{Hom}(M, Q)\big)$$

of T-U-bimodules, where ${}_R\mathrm{Hom}$ denotes the group of left R-module homomorphisms.

10.8. Let R be a commutative ring, P a finitely generated projective R-module (see Exercise 10.4), and M an R-module.

(a) Exhibit an isomorphism $\varphi\colon \mathrm{Hom}_R(P, R) \otimes_R M \cong \mathrm{Hom}_R(P, M)$ of R-modules.

(b) Define the *trace* $\mathrm{End}_R P \to R$ by composing the map

$$\varphi^{-1}\colon \mathrm{End}_R P = \mathrm{Hom}_R(P, P) \to \mathrm{Hom}_R(P, R) \otimes_R P$$

from (a) (with $M = P$) with the map $\mathrm{Hom}_R(P, R) \otimes_R P \to R$ that sends $f \otimes p$ to $f(p)$. Prove that this is the usual trace map when P is free.

10.9. Let R, A be rings, and let $f\colon R \to A$ be a ring homomorphism. View A as a (left) A-module in the usual way, and as a right R-module through $a \cdot r = a \cdot f(r)$; this makes A into an A-R-bimodule. Denote by ${}_R\mathbf{Mod}$ and ${}_A\mathbf{Mod}$ the categories of R- and A-modules, respectively. Let $F\colon {}_R\mathbf{Mod} \to {}_A\mathbf{Mod}$ be the functor $A \otimes_R -$, and $G\colon {}_A\mathbf{Mod} \to {}_R\mathbf{Mod}$ the functor that turns an A-module M into an R-module by $rm = f(r)m$. Prove that (F, G) is an adjoint pair, as defined before Exercise 7.25.

10.10. Let R be a commutative ring, and R' a commutative R-algebra. For an R-module M, we write M' for the R'-module $R' \otimes_R M$.

(a) Prove: $M' \otimes_{R'} N' \cong (M \otimes_R N)'$ as R'-modules, for any R-modules M and N.

(b) Prove that M' is projective as an R'-module if M is projective as an R-module.

(c) Prove that M' is flat as an R'-module if M is flat as an R-module.

10.11. Let R be a ring and let I be an R-module. Prove that the following three properties of I are equivalent (you may use Zorn's lemma):

(i) I is injective;

(ii) for any two R-modules M' and M, and any injective map $f\colon M' \to M$ of R-modules, the map $f^*\colon {}_R\mathrm{Hom}(M, I) \to {}_R\mathrm{Hom}(M', I)$ is surjective;

(iii) for any left ideal $J \subset R$ and any R-linear map $f\colon J \to I$ there exists $x \in I$ such that for all $r \in J$ one has $f(r) = rx$;

(iv) every short exact sequence $0 \to I \to M \to M'' \to 0$ of R-modules splits.

10.12. Prove that a $\mathbf{Z}$-module is injective if and only if it is divisible (see Exercise 6.5(a)).

10.13. Let G be a group. A *G-module* is a module over the group ring $\mathbf{Z}[G]$ (see Exercise 1.25).

(a) Prove that endowing an abelian group A with a G-module structure is equivalent to giving a group homomorphism $G \to \mathrm{Aut}\, A$.

(b) For a G-module A, write $A^G = \{a \in A : \text{for all } g \in G \text{ one has } ga = a\}$. This is a subgroup of A (you do not have to check this). Is the functor from the category of G-modules to the category of abelian groups sending A to A^G exact? left exact? right exact?

(c) Construct a G-module M such that for every G-module A there is a group isomorphism ${}_G\mathrm{Hom}(M, A) \cong A^G$ that is functorial in A.

10.14. Let R be a commutative ring, and R' a commutative R-algebra. As in Exercise 10.10, we write $M' = R' \otimes_R M$ for any R-module M, and for an R-linear map f we also write f' for the R'-linear map $f \otimes 1_{R'}$.

Let P be a finitely generated projective R-module. Exhibit an R'-algebra isomorphism $\mathrm{End}_{R'}(P') \cong \mathrm{End}_R(P)'$ that composed with the map $\mathrm{trace}'\colon \mathrm{End}_R(P)' \to R'$ gives the trace over R' (see Exercise 10.8 for the trace).

10.15. Let R be a ring, M an R-module, and F the functor $- \otimes_R M$ (from the category of right R-modules to the category of abelian groups). Recall that F commutes with taking arbitrary right limits (cf. Exercise 6.14), and that M is called *flat* if F is *exact* in the sense that for every exact sequence $A \to B \to C$ of right R-modules the sequence $FA \to FB \to FC$ is also exact.

(a) Prove that M is flat if and only if F transforms every short exact sequence into a short exact sequence, and if and only if F commutes with taking kernels.

(b) Let a diagram, as defined in Exercise 7.9, be called *finite* if the sets V and E from Exercise 7.9 are finite. Prove that M is flat if and only if it commutes with taking left limits of arbitrary finite diagrams.

10.16. Denote by $(-)^\infty$ a countably infinite direct product. Prove that the natural map $(\mathbf{Z}^\infty) \otimes_{\mathbf{Z}} \mathbf{Q} \to \mathbf{Q}^\infty$ is not an isomorphism. Deduce that $- \otimes_{\mathbf{Z}} \mathbf{Q}$ does not commute with taking arbitrary left limits, although $\mathbf{Q}$ is a flat $\mathbf{Z}$-module.

10.17. Let R, M, and F be as in Exercise 10.14. We call M *faithfully flat* if M is flat and for every non-zero right R-module N one has $FN \neq 0$.

(a) Prove: if M is free and non-zero, then M is faithfully flat.

(b) Suppose that M is faithfully flat, and let f be a morphism of right R-modules. Prove: f is injective if and only if $F(f)$ is injective; f is surjective if and only if $F(f)$ is surjective; and $f = 0$ if and only if $F(f) = 0$.

(c) Suppose that M is faithfully flat. Prove that a sequence $A \to B \to C$ of right R-modules is exact if and only if the sequence $FA \to FB \to FB$ of abelian groups is exact.

10.18. Let R be a ring. An R-module M is called *finitely presented* if there is an exact sequence $R^m \to R^n \to M \to 0$ of R-modules, with n, m non-negative integers.

(a) Prove that every finitely generated projective R-module is finitely presented.

(b) Suppose that R is left noetherian (see Exercise 8.12), and let M be an R-module. Prove: M is finitely presented if and only if it is finitely generated.

10.19. Let R, R', and the notation $'$ be as in Exercise 10.14. Suppose that R' is flat over R.

(a) Prove: if M is a finitely presented R-module, and N is any R-module, then one has $\mathrm{Hom}_R(M, N)' \cong \mathrm{Hom}_{R'}(M', N')$ as R'-modules. (*Hint.* First do the case in which M is free.)

(b) Suppose that R' is faithfully flat over R (see Exercise 10.17), and let M be an R-module. Prove: M is finitely generated as an R-module if and only if M' is finitely generated as an R'-module. Prove also: M is finitely generated and projective as an R-module if and only if M' is finitely generated and projective as an R'-module.

11. Finite étale algebras

For this section let R be a commutative ring.

We return to our observation that for any R-algebra A and R-module M, $A \otimes_R M$ is an A-module, and that this defines a functor

$$F \colon {}_R\mathbf{Module} \to {}_A\mathbf{Module}$$

called an "extension of scalars" functor. We often denote the extension of scalars to a module by subscripting it appropriately:

$$F(M) = M_A = A \otimes_R M.$$

Many useful properties held by M_R which we will consider will hold also for M_A. In particular, this will be true of the properties of being projective, flat, and free, or more more specifically, free of a given rank. As an example, we shall show the last claim.

Example. Let M_R be a free R-module, and let S be such that $M_R \cong R^{(S)}$. Then

$$\begin{aligned} M_A &= A \otimes_R M_R \\ &\cong_R A \otimes_R R^{(S)} \\ &\cong_R (A \otimes_R R)^{(S)} \\ &\cong_R A^{(S)}, \end{aligned}$$

as we have already observed in the previous section on the properties of tensor product. and a simple check shows that the isomorphisms hold when viewed as A-modules rather than R-modules. It is also clear from this that the rank must be the same.

If R is a field, and M is a finite-dimensional R-module (in other words a finite-dimensional vector space over R), we know from linear algebra that we can describe any R-linear endomorphism from M to M by a matrix (provided that we are given a basis for M). In fact, this description does not in general require that R be a field, but only that it be a commutative ring and that M is a free R-module of finite rank n. If we are equipped with a basis for M, we may write any endomorphism $\varepsilon\colon M \to M$ as a matrix E. In other words, if $\mathrm{End}_R(M)$ is the ring of R-linear endomorphisms in M and $M(n, R)$ is the ring of $n \times n$-matrices over R, then

$$\mathrm{End}_R(M) \overset{\Phi}{\cong} M(n, R),$$

where Φ is a ring isomorphism depending only on the choice of basis for M.

Many of the our notions from linear algebra carry over into this new ring. Given any matrix $E \in M(n, R)$ we can define functions

$$\begin{aligned} \mathrm{Tr}\colon M(n, R) &\to R, \\ \mathrm{Det}\colon M(n, R) &\to R, \\ \Xi\colon M(n, R) &\to R[t], \end{aligned}$$

(the trace, determinant, and characteristic polynomial respectively) by constructions that are completely analogous to their counterparts in linear algebra. For example, if the entries of a matrix E are indexed e_{ij}, we define Tr by

$$\mathrm{Tr}(E) = \sum_{i=1}^{n} e_{ii}.$$

We would like these functions to be defined on $\mathrm{End}_R(M)$, and the above isomorphism Φ allows us to do this, provided that these functions are independent of the choice of basis. In other words, if ε is an R-linear endomorphism in M, we define $\mathrm{Tr}(\varepsilon)$ by

$$\mathrm{Tr}(\varepsilon) = \mathrm{Tr}(\Phi(\varepsilon)),$$

which is well-defined as long as the trace is independent of the choice of a basis. It is an exercise from linear algebra to show that this is true (as is the corresponding statement for the determinant, etc.). In short, one can show that $\mathrm{Tr}(BC) = \mathrm{Tr}(CB)$, and from this it follows that $\mathrm{Tr}(UAU^{-1}) = \mathrm{Tr}(A)$, by setting $U = B$ and $AU^{-1} = C$. Furthermore, one can show that any change of basis is accomplished by conjugating by an invertible matrix U, and this completes the proof that the trace is well-defined regardless of a choice of basis. There is a cleaner way to show this, however, without explicitly computing with the linear algebra, except to invoke the existence of a basis.

For the remainder of this section, let A be a commutative R-algebra, which when viewed as an R-module is free of finite rank n. (Note: much of what will be said only needs that A is projective and finitely generated, but we will stick with the stronger condition, as this suffices for our considerations.)

The trace map Tr can now be viewed as a map from A to R, denoted by $\mathrm{Tr}_{A/R}$, by defining $\mathrm{Tr}_{A/R}(a)$ to be $\mathrm{Tr}(\mu_a)$, where μ_a is the R-linear endomorphism in A that maps x to ax.

Example. Let R be the ring $\mathbf{Q}$, and A be the $\mathbf{Q}$-algebra $\mathbf{Q}(\sqrt{5})$. Then A is free of rank 2 over R, and every element is of the form $a = r + s\sqrt{5}$, where $r, s \in \mathbf{Q}$. The matrix associated with the map $(x \mapsto ax)$ is

$$\begin{bmatrix} r & 5s \\ s & r \end{bmatrix},$$

and its trace is $2r$. Therefore

$$\mathrm{Tr}_{\mathbf{Q}(\sqrt{5})/\mathbf{Q}}(r + s\sqrt{5}) = 2r.$$

A simple check reveals that $\mathrm{Tr}_{A/R}$ is always an R-linear morphism from A to R (as R-modules). Therefore, we may regard $\mathrm{Tr}_{A/R}$ as an element of $\mathrm{Hom}_R(A, R)$ which is the R-module dual to A. That $\mathrm{Hom}_R(A, R)$ is also free of rank n can be seen quite easily, since if $A \cong R^n$, then

$$\mathrm{Hom}_R(A, R) \cong \mathrm{Hom}_R(R^n, R) \cong \big(\mathrm{Hom}_R(R, R)\big)^n \cong R^n.$$

(Note: if we relaxed our condition that A be free of finite rank as an R-module, to say that it is finitely generated and projective, then it would again be true that $\mathrm{Hom}_R(A, R)$ is finitely generated and projective, but A and $\mathrm{Hom}_R(A, R)$ might *not* be isomorphic as R-modules.)

Finally, we can construct a module morphism $\Phi\colon A \to \mathrm{Hom}_R(A, R)$ which maps

$$a \overset{\Phi}{\mapsto} (a\mathrm{Tr}_{A/R}) = \big(b \overset{\Phi(a)}{\mapsto} \mathrm{Tr}_{A/R}(ba)\big).$$

We are now ready to define the concept of finite étale.

Definition. *Let R be a commutative ring, A a commutative R-algebra which is free of rank n as an R-module. We say A is* **finite étale** *if the map Φ described above is an isomorphism.*

Note that the map Φ is both R-linear and A-linear. Therefore, if A is finite étale, A and $\mathrm{Hom}_R(A, R)$ are isomorphic both as R-modules and as A-modules. The map Φ described above sends the identity 1 to the element Tr. In other words, considered as an A-module, $\mathrm{Hom}_R(A, R)$ must be free of rank 1, with the trace map Tr as a basis element. There is a notion of general étale which is not restricted by the condition of finiteness, but it will not be used in this course, and we may free our minds of its existence.

We leave the slight generalization for projective modules to the reader.

Example. Let us take again the case $A = \mathbf{Q}(\sqrt{5}), R = \mathbf{Q}$. We consider the action of the map Φ on the basis elements $1, \sqrt{5}$ of A.

We examine what our map $\Phi\colon \mathbf{Q}(\sqrt{5}) \to \mathrm{Hom}_{\mathbf{Q}}(\mathbf{Q}(\sqrt{5}), \mathbf{Q})$ does on the basis elements of $\mathbf{Q}(\sqrt{5})$, 1 and $\sqrt{5}$.

$$\begin{array}{ccc} 1 & \mapsto & \begin{cases} 1 & \mapsto & 2 \\ \sqrt{5} & \mapsto & 0 \end{cases} \\ \sqrt{5} & \mapsto & \begin{cases} 1 & \mapsto & 0 \\ \sqrt{5} & \mapsto & 10 \end{cases} \end{array}$$

Thus if we use as our basis elements for $\mathrm{Hom}_{\mathbf{Q}}(\mathbf{Q}(\sqrt{5}), \mathbf{Q})$ the two maps ϕ_1, ϕ_2 where $\phi_1(1) = 1$, $\phi_1(\sqrt{5}) = 0$, $\phi_2(1) = 0$, $\phi_2(\sqrt{5}) = 1$, the transformation matrix associated to Φ with respect to this basis is

$$\begin{bmatrix} 2 & 0 \\ 0 & 10 \end{bmatrix},$$

which is nonsingular. Hence $\mathbf{Q}(\sqrt{5})$ is finite étale over $\mathbf{Q}$.

We note that the same computation as above shows that $\mathbf{Z}[\sqrt{5}]$ is *not* finite étale over $\mathbf{Z}$, since 20 is not invertible in $\mathbf{Z}$.

Let L, K be fields, L a finite extension of K of degree n. Then L is finite étale over K if and only if the trace map $\mathrm{Tr}\colon L \to K$ is not the zero function, as we will show in the next lemma. In other words, we have a sequence of implications

$$\begin{aligned} \operatorname{char} K = 0 &\Rightarrow n \not\equiv 0 \bmod \operatorname{char} K \\ &\Rightarrow \mathrm{Tr}(1) \neq 0 \\ &\Rightarrow L \text{ is finite étale over } K. \end{aligned}$$

Lemma. *Let L, K be fields, and L a finite extension of K of degree n. Then L is finite étale over K if and only if the trace map $\mathrm{Tr}\colon L \to K$ is not the zero function.*

Proof. Recall from previous observations that $\mathrm{Hom}_L(L, K)$ is an L-vector space. Since the dimension of $\mathrm{Hom}_K(L, K)$ as a K-vector space is n, we know from linear algebra that

$\dim_L(\mathrm{Hom}_K(L, K)) = 1$. Hence Φ is an isomorphism exactly when the trace map spans the one-dimensional space, which happens exactly when it is nonzero.

As an immediately corollary, we see that every finite extension of a field of characteristic zero is finite étale. For vocabulary's sake, we introduce another term.

Definition. *Let L be a finite extension of K. We say L is* **separable** *over K if L is finite étale over K.*

The reader may already be familiar with the notion of separability in connection with the number of possible extensions of embeddings of K into an algebraically closed field $\bar{K}$. It turns out that the definitions are equivalent, but this will be shown later.

It may sound to the reader as if *all* field extensions are finite étale; we now give a counterexample. Let $K = \mathbf{F}_p(t)$, the field of rational functions over $\mathbf{F}_p$, and $L = K(\sqrt[p]{t})$, where we adjoin the p-th root of the variable t. Then this extension is inseparable, since the trace map is identically zero.

We collect here some properties of finite étale algebras which will be useful in the sequel.

(i) Viewed as an R-algebra, R itself is finite étale over R. That R is free, of rank 1, is obvious. Furthermore, the above map Φ (from the definition prior to the last) sends an element r to its corresponding element in its dual, $\mathrm{Hom}_R(R, R)$, and this is clearly an isomorphism.

(ii) If A and B are both finite étale over R, so is $A \times B$ (as an R-algebra), and conversely. We will not give a full explanation, but will sketch the reasons here. That $A\times B$ is free of finite rank as an R-module is again clear. The trace over $A \times B$ of an element $(a, b) \in A \times B$ is

$$\mathrm{Tr}_{(A\times B)/R}(a, b) = \mathrm{Tr}_{A/R}(a) + \mathrm{Tr}_{B/R}(b).$$

We see this by writing the matrix representation of the maps μ_a, μ_b by picking some appropriate basis; i.e. we have

$$\begin{bmatrix} M & 0 \\ 0 & N \end{bmatrix}.$$

Also, as R-modules, we have already seen that

$$\mathrm{Hom}_R(A \times B, R) \cong \mathrm{Hom}_R(A, R) \times \mathrm{Hom}_R(B, R).$$

(iii) If we give the R-module R^n an R-algebra structure through componentwise multiplication, it is also finite étale. This result follows directly from the last two. Note that this is true even if $n = 0$, meaning that the zero ring is trivially finite étale over R.

An R-algebra A is called *totally split* if $A \cong R^n$, *as R-algebras* (not as R-modules!), for some $n < \infty$.

(iv) We have already discussed the extension of scalars functor, and we shall now introduce notation to accompany it: If R' is a commutative R-algebra, we denote $R' \otimes_R A$ by A'. So if A is finite étale over R, then A' is finite étale over R'. To prove this, we simply check that the change of scalars functor preserves all the relevant properties involved. As an R-algebra, R' is equipped with a homomorphism $f\colon R \to R'$. Then for all $a \in A$ and its corresponding $a' = a \otimes_R 1 \in A'$, we have

$$f(\mathrm{Tr}_{A/R}(a)) = \mathrm{Tr}_{A'/R'}(a').$$

In other words, if $M \cong R^n$ as an R-module, the following diagrams commute:

$$\begin{array}{ccc} \mathrm{Hom}_R(M, M) & \longrightarrow & \mathrm{Hom}_{R'}(M', M') \\ {\scriptstyle\text{Trace}}\big\downarrow & & \big\downarrow{\scriptstyle\text{Trace}} \\ R & \xrightarrow{f} & R' \end{array}$$

$$\begin{array}{ccc} A & \xrightarrow{R' \otimes_R -} & A' \\ {\scriptstyle\Phi_A}\big\downarrow & & {\scriptstyle\Phi_{A'}}\big\downarrow \\ \mathrm{Hom}_R(A, R) & \longrightarrow & \mathrm{Hom}_{R'}(A', R'). \end{array}$$

In fact, in the top diagram, the right trace map is equal to the tensor product of the left map with f, modulo the necessary identifications. The proof that these diagrams commute is straightforward.

(v) Moreover, if R' is free and non-zero as an R-module, we have the following result:

$$A \text{ is finite étale over } R \Leftrightarrow A' \text{ is finite étale over } R'.$$

Namely, if R' is free, the converse of (iv) also holds. To show this, we need to show that the map $\Phi_A\colon A \to \mathrm{Hom}_R(A, R)$ is an isomorphism if and only if the induced map $\Phi_{A'}\colon A' \to \mathrm{Hom}'_R(A', R')$ is an isomorphism. Since R' is free as an R-module, we have a module-isomorphism $R' \cong_R R^{(S)}$. (Note that $S \neq \emptyset$, since R' is assumed to be nonzero.) Then $A' = A^{(S)}$, as before, and moreover

$$\mathrm{Hom}_{R'}(A', R') = (\mathrm{Hom}_R(A, R))^{(S)}.$$

But then $\Phi_{A'} = \Phi_A^{(S)}$, so clearly Φ_A is an isomorphism if and only if $\Phi_{A'}$ is an isomorphism.

This already leads us to some interesting results which we could have proved earlier, but from a different perspective.

Example. Let $f \in \mathbf{Q}[x]$ an irreducible monic polynomial. Then

$$\mathbf{Q}[x]/(f) = \mathbf{Q}[x]/f\mathbf{Q}[x]$$

is a finite field extension of $\mathbf{Q}$, and since the characteristic of $\mathbf{Q}$ is zero, this is finite étale. Following the above notation, let $A = \mathbf{Q}[x]/(f)$, $R = \mathbf{Q}$, $R' = \mathbf{C}$. Then we know that $A' = A \otimes_{\mathbf{Q}} \mathbf{C}$ is finite étale over $\mathbf{C}$.

Let us examine more closely the tensor product $A \otimes_{\mathbf{Q}} \mathbf{C}$. This is precisely $\mathbf{C}[x]/f\mathbf{C}[x]$. To see this, we use the exactness properties of the tensor product. We have the exact sequence

$$\mathbf{Q}[x] \xrightarrow{\cdot f} \mathbf{Q}[x] \xrightarrow{\pi} \mathbf{Q}[x]/(f) \to 0,$$

which we can tensor with $\mathbf{C}$. Since $\mathbf{C} \otimes_{\mathbf{Q}} \mathbf{Q}[x] = \mathbf{C}[x]$, we have

$$\mathbf{C}[x] \xrightarrow{\cdot f} \mathbf{C}[x] \to A \otimes_{\mathbf{Q}} \mathbf{C} \to 0$$

is an exact sequence. Notice, however, that we know the structure of $A \otimes_{\mathbf{Q}} \mathbf{C}$ explicitly. Since f is irreducible, we know that each factor $(x-a)$ in the factorization of f in $\mathbf{C}[x]$ is of multiplicity one, i.e.

$$f = \prod_{i=1}^{t} (x - a_i)^{m(i)} = \prod_{i=1}^{t} (x - a_i),$$

where the roots a_i are distinct. So the Chinese Remainder Theorem says

$$\begin{aligned} \mathbf{C}[x]/f\mathbf{C}[x] &\cong \prod_{a_i} \mathbf{C}[x]/(x-a)\mathbf{C}[x] \\ &\cong \mathbf{C}^{\deg(f)}, \end{aligned}$$

as $\mathbf{C}$-algebras. So $A \otimes_{\mathbf{Q}} \mathbf{C}$ is totally split over $\mathbf{C}$, so certainly finite étale.

In fact, we can describe the category of totally split K-algebras, $\mathbf{TotallySplitAlg}_K$, with K-algebra morphisms between them. We now describe in detail these morphisms. The objects in this category are the K-algebras isomorphic (as K-algebras) to K^n, for some $n \in \mathbf{Z}_{\geq 0}$. Given two such objects K^m, K^n, a morphism Φ between them is equivalent to giving n morphisms $\phi_i \colon K^m \to K$, since the K^n is a direct product. Hence to count the number of possible morphisms from K^m to K^n, it suffices to count the number of possible K-algebra homomorphisms from K^m to K. Each of them is certainly a surjective map, since it is K-linear. But the kernel must be a maximal ideal, since the image is a field.

But we have already described all maximal ideals of a product of rings, and since each component is a field, we see that the only possible K-linear maps are the projections. Hence we have m^n possible morphisms.

We note that this is the reverse of what happens in the category of sets, where there are n^m possible morphisms from a set of size m to a set of size n. We have the following theorem.

Theorem. *For any field K, the category $\mathbf{TotallySplitAlg}_K$ of totally split K-algebras is anti-equivalent to the category $\mathbf{FiniteSet}$ of finite sets.*

The theorem actually holds in more generality, for the category of totally split R-algebras, where R commutative and connected. In other words, we can define contravariant functors

$$\begin{aligned} F\colon \mathbf{TotallySplitAlg}_R &\to \mathbf{FiniteSet} \\ A &\mapsto \{R\text{-algebra homomorphisms } \Phi\colon A \to R\}, \\ \text{and} \qquad G\colon \mathbf{FiniteSet} &\to \mathbf{TotallySplitAlg}_R \\ S &\mapsto R^S = \{\text{maps } f\colon S \to R\}. \end{aligned}$$

We can exhibit isomorphisms of the functors $1_{\mathbf{TotallySplitAlg}_R} \to GF$ and $1_{\mathbf{FiniteSet}} \to FG$. For instance, an isomorphism between A and $GF(A) = \{\text{maps: } \mathbf{Algebra}_K(A, K) \to K\}$ is given by taking the element a to $(f \mapsto f(a))$, the evaluation homomorphism.

Before we go on to the discussion of more general finite étale K-algebras, we note one more fact.

Theorem. *Let K be a field, A a totally split K-algebra, and $B \subset A$ a sub-K-algebra. Then B is also totally split.*

Proof. Note first that any sub-K-algebra is by definition also a subring, and hence contains the unit element. Without loss of generality, we consider $A = K^n$ for some finite $n \in \mathbf{Z}_{\geq 0}$. The n projection maps $K^n \to K$ restricted to B give surjective K-algebra maps $B \to K$. (The map is surjective since B contains the unit, and the projections π_i are K-linear.) The kernels $M_1, M_2, \ldots, M_n$ of the π_i are maximal ideals since the image is a field in each case, and the intersection is trivial, since if $\pi_i(x) = 0$ for each projection, then $x = 0 \in K^n$.

Note that we do *not* know that the M_i are pairwise distinct. We take some subset $M_1, M_2, \ldots, M_t$ which *are* pairwise distinct, with t maximal. Then they are pairwise coprime. We still have $\bigcap_{i=1}^t M_i = 0$, so by the Chinese Remainder Theorem,

$$B \cong B/(\bigcap_{i=1}^{t} M_i) \cong \prod_{i=1}^{t} B/M_i \cong K^t.$$

Thus B is also totally split, as required.

Note that this theorem is *not* a consequence of the previous theorem about the category of totally split K-algebras. This proof depends upon the fact that K is a field, whereas (as we have already noted) the anti-equivalence of categories holds in more generality.

Example. We give a counterexample to the above theorem in the case of an R-algebra, R not a field. Consider

$$A = \{(a, b) \in \mathbf{Z} \times \mathbf{Z} : n \equiv m \pmod 2\} \subset \mathbf{Z} \times \mathbf{Z}.$$

Then A is a sub-$\mathbf{Z}$-algebra of a totally split $\mathbf{Z}$-algebra, but is not totally split over $\mathbf{Z}$. In fact, it is not even finite étale over $\mathbf{Z}$.

There is a theorem (which we will not prove) which states that a $\mathbf{Z}$-algebra is finite étale if and only if it is totally split. Recall that this is not the case for the rationals $\mathbf{Q}$ or for the reals $\mathbf{R}$, but is the case for the complex numbers $\mathbf{C}$; any finite field extension of $\mathbf{Q}$ is finite étale but certainly does not need to be totally split.

We now make a definition which will allow us to classify certain K-algebras which are particularly simple in structure.

Definition. *A field $\bar{K}$ is called* **separably closed** *if it has no finite separable field extension, except $\bar{K}$ itself.*

Theorem. *Let K be a field, and A a K-algebra. Then A is finite étale if and only if A is isomorphic (as a K-algebra) to the finite product $\prod L_i$ for certain finite separable field extensions $L_i \supset K$. In particular, if $\bar{K}$ is separably closed, then A is finite étale if and only if A is totally split.*

Proof. The "if" direction is clear. In the "only if" direction, let A be a finite K-algebra, so A is isomorphic to a finite product of $A/M^{n(M)}$, for maximal ideals $M \subset A$. So A is finite étale if and only if each $A/M^{n(M)}$ is finite étale.

We claim $M \cong M^{n(M)}$. Let $\tilde{A} = A/M^{n(M)}$, and $\tilde{M} = M/M^{n(M)}$. Let a be an element of $\tilde{M}$. Then the endomorphism in $\tilde{A}$ defined by $x \mapsto ax$ is clearly nilpotent. From linear algebra we know that the trace of any nilpotent endomorphism is 0. So the image of the trace map $\mathrm{Tr}_{\tilde{A}/K}$ restricted to $\tilde{M}$ is $\{0\}$. We know that $\tilde{A}$ is finite étale, so the mapping from $\tilde{A}$ to $\mathrm{Hom}_K(\tilde{A}, K)$ defined by

$$b \mapsto (c \mapsto \mathrm{Tr}(bc))$$

is an isomorphism. So every $f \in \mathrm{Hom}_K(\tilde{A}, K)$ "kills" $\tilde{M}$. This means that the kernel $\tilde{M}$ must be trivial. So since $\tilde{M}$ is trivial, we know that $M = M^{n(M)}$, proving the claim.

So each $\tilde{A}$ is A/M, which is a field. Therefore, A is the product of fields, which must be finite separable extensions of K. In particular, if $\bar{K}$ is separably closed, each $L_i = \bar{K}$, so A is finite étale if and only if A is totally split.

We mention here a few facts about algebraic closures of fields, with which we assume the reader is already familiar.

Definition. *Let $\bar{K}$ be a field. We say $\bar{K}$ is* **algebraically closed** *if the only field extension L of $\bar{K}$ which is finite is $L = \bar{K}$ itself.*

Definition. *A field extension $\bar{K}$ of K is an* **algebraic closure** *of K if these conditions are satisfied:*

(i) $\bar{K} \supset K$ is algebraic, i.e. every element $\beta \in \bar{K}$ is the root of some polynomial in $K[x]$.

(ii) $\bar{K}$ is algebraically closed.

So the definition of an algebraically closed field matches the definition of a separably closed field identically, but without the word “separable.” Thus any algebraically closed field is automatically separably closed. The condition of separable closure is therefore weaker than algebraic closure, but it turns out that it is not much weaker. For example, the conditions are the same if $\bar{K}$ has characteristic 0.

We mention two useful facts, the proofs of which can be found in Lang (Ch. V §2).

Theorem. *Every field K has an algebraic closure.*

Theorem. *Let K, L be fields, and $\sigma\colon K \to L$ a field isomorphism. Suppose $K \subset \bar{K}$, an algebraic closure of K. Then if $\bar{L}$ is an algebraic closure of L, there exists an isomorphism $\tau\colon \bar{K} \to L$ that extends σ.*

It is important to note in the latter fact that the isomorphism τ is usually *not* unique. Hence, although some authors refer to *the* algebraic closure of a field K, one should in fact refer to *an* algebraic closure. (Similarly, one speaks of *a* field of order $q = p^n$, where p is prime, since the isomorphism between fields of order q is not canonical if $n > 1$.)

We also assume that the reader is familiar with some elementary properties of algebraic closures. For instance, given $L \supset K$ an algebraic extension and $\bar{L}$ an algebraic closure of L, then $\bar{L}$ is also an algebraic closure of K.

We indulge next in a casual and motivational comparison of the classical and modern approaches to Galois theory. In all current textbooks, Galois theory is studied using finite separable field extensions L of a given base field K. Our approach follows that of the Grothendieck formulation, in which the objects under consideration are finite étale K-algebras A. We now consider the relation between the two perspectives.

Recall that we have shown earlier on that all finite étale K-algebras are a finite product of finite field extensions of K, i.e. given a finite étale K-algebra A,

$$A \cong_K \prod_{i=1}^{t} L_i,$$

where each L_i is a finite separable extension.

One can see from this decomposition of finite étale K-algebras that there is only a marginal difference, so to speak, between the objects under study in the modern formulation and those in the classical formulation. We have already emphasized that products of rings are rarely more interesting than the individual rings. One could still ask, nevertheless, why we have decided to consider such products of field extensions in the first place—especially if they are no more interesting than their components. It turns out that finite étale algebras satisfy some crucial properties which finite field extensions do not. Recall that many "good" properties of an R-algebra A were preserved under base changes, as has been observed previously. We refer the reader back to the previous section. For instance, we already know that for a finite extension L of a field K and a finite-dimensional algebra A over K, A is finite étale over K if and only if $A' = A \otimes_K L$ is finite étale over L. Hence the property of being finite étale is preserved under base changes. We have the following diagram.

$$\begin{tikzcd} & A \otimes_K L & \\ A \arrow[ur] & & L \arrow[ul, "\text{f.é.}"'] \\ & K \arrow[ul, "\text{f.é.}"] \arrow[ur] & \end{tikzcd}$$

On the other hand, what is the situation in classical Galois theory? Perhaps the analogous statement is true. Namely, given a field K and two field extensions L and L', is it true that L is finite separable over K if and only if $L \otimes_K L'$ is finite separable over L'?

$$\begin{tikzcd} & L \otimes_k L' & \\ L \arrow[ur] & & L' \arrow[ul, "\text{finite separable}"'] \\ & K \arrow[ul, "\text{finite separable}"] \arrow[ur] & \end{tikzcd}$$

The question may at first seem odd, since we have *defined* an extension to be separable if it is finite étale. However, it is *not* necessarily the case that the object $L \otimes_K L'$ is again a field, so in fact the diagram above does *not* hold, as the following example illustrates.

Example. Let $K = \mathbf{Q}$, $L = \mathbf{Q}(\sqrt{5})$, $L' = \mathbf{R}$. Then $L \otimes_K L' = \mathbf{Q}(\sqrt{5}) \otimes_{\mathbf{Q}} \mathbf{R} = \mathbf{R}[x]/(x^2 - 5) = \mathbf{R} \times \mathbf{R}$. This is clearly not a field, so we cannot even hope that it is a finite separable extension.

So we see that in order to preserve the "good" qualities of extensions under base changes, it is necessary to expand the universe under our consideration.

We now proceed to some further results which make it clear why it is easier to talk about finite étale algebras than finite separable extensions. Recall that a separably closed field is one for which there does not exist any finite separable extension except itself. Since any field has an algebraic closure, any field is contained in a separably closed field.

Theorem. *Let K be a field, $\bar{K}$ a separably closed field extension of K. Let A be a K-algebra, and define $A_{\bar{K}} = A \otimes_K \bar{K}$. Then A is finite étale over K if and only if $A_{\bar{K}}$ is totally split over $\bar{K}$.*

Proof. This follows from our previous results as the following sketch shows. We know already that A is finite étale over K if and only if $A_{\bar{K}}$ is finite étale over $\bar{K}$. By previous results, $A_{\bar{K}}$ is a finite product of finite separable extensions L_i of $\bar{K}$. Since $\bar{K}$ is separably closed, each such L_i is $\bar{K}$ itself. Hence $A_{\bar{K}} = \bar{K}^n$ for some finite n and hence is totally split over $\bar{K}$.

Corollary. *Let A be a finite étale algebra over a field K, and $B \subset A$ a sub-K-algebra. Then B is also finite étale over K.*

Proof. Again, we give a sketch of the proof. The inclusion map $B \hookrightarrow A$ gives an inclusion $B_{\bar{K}} \hookrightarrow A_{\bar{K}}$. That is, given the exact sequence

$$0 \to B \to A,$$

we use the fact that the tensor product $- \otimes \bar{K}$ is an exact functor (i.e. a free module is flat) and see that

$$0 \to B_{\bar{K}} \to A_{\bar{K}}$$

is also an exact sequence. By the previous theorem, $A_{\bar{K}}$ is totally split over $\bar{K}$, and (being a sub-$\bar{K}$-algebra) $B_{\bar{K}}$ is also totally split over $\bar{K}$, and hence B is finite étale over K.

Note that if L is a field extension of K and A is an L-algebra, we can view A as a K-algebra in a natural way. Namely, if A is an L-algebra, then there exists a homomorphism $\Phi: L \to A$. Since we also have the inclusion map $K \hookrightarrow L$, we compose the two maps to get a homomorphism $\Phi': K \to A$, and this yields the K-algebra structure on A.

Theorem. *Let $L \supset K$ be a finite separable extension of fields. Let A be a finite étale L-algebra. Then A is also finite étale over K.*

Proof. We give a sketch of the proof. We tensor everything with $\bar{K}$, where $\bar{K}$ is a separably closed field extension of K. That is, we have two sequences

$$\begin{array}{ccccc} K & \longrightarrow & L & \overset{\text{f.é.}}{\longrightarrow} & A \\ \bar{K} & \longrightarrow & L_{\bar{K}} & \overset{\text{f.é.}}{\longrightarrow} & A_{\bar{K}} \end{array}$$

and since $\bar{K}$ is separably closed, $L_{\bar{K}} = \bar{K}^n$, since $L_{\bar{K}}$ is finite étale. It can be shown routinely that any algebra A over a product of rings $R_1 \times R_2$ is actually a product of algebras $A_1 \times A_2$. The proof follows that for the case of modules over a product of rings. If A is an algebra over $R = R_1 \times R_2$, then A is finite étale if and only if A_1 is finite étale over R_1, and A_2 is finite étale over R_2. We leave this for the reader. Since $A_{\bar{K}}$ above is finite étale over $\bar{K}^n$, this means $A_{\bar{K}} = B_1 \times B_2 \times \cdots \times B_n$, where each B_i is finite étale over the i-th component ring $\bar{K}$ of $\bar{K}^n$. Each B_i is totally split over $\bar{K}$, since $\bar{K}$ separably closed. So $A_{\bar{K}}$ is a finite product of copies of $\bar{K}$, and hence totally split over $\bar{K}$. Hence by our previous theorem, A is finite étale over K.

We can now begin to compare our results with results which may appear more familiar to the reader acquainted with classical Galois theory. First we introduce the notion of a derivative.

Definition. *Let R be a commutative ring, and let $R[x]$ be the ring of polynomials. For any polynomial $f \in R[x]$ such that*

$$f = \sum_{i=0}^{t} r_i x^i,$$

we define the **derivative** *of f to be the polynomial $f' \in R[x]$, where*

$$f' = \sum_{i=1}^{t} i r_i x^{i-1}.$$

Theorem. *Let k be a field, and let $f \in K[x]$ be a non-zero polynomial. Then $K[x]/fK[x]$ is finite étale over K if and only if* $\gcd(f, f') = 1$, *or equivalently, $fK[x] + f'K[x] = K[x]$.*

In fact, this statement is valid not just in the case of fields, but for any commutative ring.

Before proving the theorem, we remind the reader of a useful fact. Note that we can always normalize our polynomials to be monic; this allows us to define the greatest common divisor of two polynomials *uniquely.* Recall that f is divisible by g if and only if f is contained in the ideal (g) generated by g. Hence we define the greatest common divisor of two polynomials f, g to be the unique monic generator of the ideal (f, g). (We assume the ideal is non-zero.)

Lemma. *Let K be a field, $L \supset K$ a finite extension, and $f \in K[x]$ a non-zero polynomial. Then*
$$(\gcd(f, f') \in K[x]) = (\gcd(f, f') \in L[x]).$$

Proof. The greatest common divisor of f and f' in $K[x]$ is characterized by being the monic polynomial with the property
$$fK[x] + f'K[x] = \gcd(f, f')K[x].$$

If this holds for K then upon multiplying this identity by $L[x]$ one sees that it also holds for L. This proves the lemma.

We now prove the theorem.

Proof. Let $\bar{K}$ be algebraically closed, and $K \subset \bar{K}$. We now reduce the proof of the theorem to the corresponding result over $\bar{K}$. We already know from a theorem earlier on that $K[x]/(f)$ is finite étale if and only if $\bar{K}[x]/(f)$ is finite étale (and since $\bar{K}$ separably closed, totally split) over $\bar{K}$. Since $\bar{K}$ is algebraically closed, f splits linearly in $\bar{K}[x]$, i.e.
$$f = \prod_{i}^{<\infty}(x - a_i)^{n_i}.$$

Hence $\bar{K}[x]/(f)$ is totally split if and only if each $\bar{K}[x]/(x - a_i)^{n_i}$ is finite étale over $\bar{K}$. This is true if and only if each factor is linear, i.e. $n_i = 1$, for all i. Hence f can have no double roots in $\bar{K}$, and hence $\bar{K}[x]/(f)$ is finite étale over $\bar{K}$ if and only if $\gcd(f, f') = 1$.

We can now define the notion of separability for an element $\alpha \in L$, a finite extension of K. Let $f_K^\alpha = \mathrm{Irr}(\alpha, K)$ be the unique monic irreducible polynomial f of α over K. Then we know that
$$K(\alpha) \cong_K K[x]/(f).$$

Definition. *Let α, L, K and f be as above. Then α is* **separable** *over K if $K(\alpha)$ is separable over K, or equivalently, if f and f' are coprime.*

Under what conditions is an element α *not* separable over a field K? When we parse down the requirements, we see that in fact, this occurrence is relatively rare. (A field in which this never happens—i.e. every finite extension is separable—is called *perfect.*) Suppose α is not separable over K. Then f and f' are not coprime, but since f is an irreducible polynomial, this immediately implies that f divides f'. But since f' is of degree strictly less than f, this means that $f' = 0$. Recall that the derivative of any non-zero polynomial $f = \sum_{i=0}^{n} a_i x^i$ is defined to be $f' = \sum_{i=1}^{n} i a_i x^{i-1}$. Since at least one of the a_i is non-zero, if f' is zero, this means that the characteristic of K is positive; let $\mathrm{char}(K) =$

$p > 0$. Since each $ia_i = 0$, each i is a multiple of p, and we have that $f = g(x^p)$, for some $g \in K[x]$. Since f is irreducible, this implies that the Frobenius map $F\colon K \to K$ which sends $a \mapsto a^p$ can not be surjective. Otherwise, f would be reducible, namely

$$f = \sum_{i=1}^{n} a_i x^{ip} = \sum_{i=1}^{n} b_i^p x^{ip} = (\sum_{i=1}^{n} b_i x^i)^p.$$

(Recall that $(a+b)^p = a^p + b^p$ in characteristic p.) In particular, since the Frobenius map is always surjective for finite fields, we have just shown that any finite extension of a finite field is separable.

We are now in a position to prove the equivalence of the conditions of separability.

Theorem. *Let L be a finite field extension of K. The following are equivalent:*
(i) L is separable over K (i.e. it is finite étale).
(ii) Every element $\alpha \in L$ is separable over K.
(iii) $L = K(\alpha_1, \alpha_2, \ldots, \alpha_t)$, where each α_i is separable over K.

Proof. (i) $\Rightarrow$ (ii). Let α be any element in L. We have a tower of extensions $K \subset K(\alpha) \subset L$, where $K(\alpha)$ is a sub-K-algebra of L. We have already seen that a subalgebra of a finite étale algebra over K is also finite étale. Hence α is separable.

(ii) $\Rightarrow$ (iii). Since L is a finite extension, and in particular a finite dimensional vector space over K, we can certainly write $L = K(\alpha_1, \alpha_2, \ldots, \alpha_t)$. Since each element in L is separable, so is each α_i.

(iii) $\Rightarrow$ (i). Suppose $L = K(\alpha_1, \alpha_2, \ldots, \alpha_t)$, and each α_i is separable over K. We can do this by induction on t. Recall that if $K \subset L \subset M$ is a tower of extensions and L is separable over K and M is separable over L, then M is also separable over K. We can apply this theorem and do a straightforward induction, provided that we know that α_t is separable over $K' = K(\alpha_1, \alpha_2, \ldots, \alpha_{t-1})$. To show this, we only need to show that the irreducible polynomial of α_t is coprime to its derivative. Hence it would suffice to show that it has no double roots. But since the irreducible polynomial $\mathrm{Irr}(\alpha_t, K')$ of α_t over K' must divide the irreducible polynomial $\mathrm{Irr}(\alpha_t, K)$ of α_t over K, since α_t is separable over K, this implies that $\mathrm{Irr}(\alpha_t, K')$ also has no double roots. Hence L is separable over K, as required.

We now consider the separable closure of a field K, which will have useful properties somewhat akin to those of algebraic closures. Let K be a field, and $\bar{K}$ an algebraic closure of K. Let K_s be the subset of $\bar{K}$ of elements which are separable over K. Then K_s is a subfield of $\bar{K}$. To show this, we must show that the sum, product, and other algebraic relatives of two elements α, β of K_s are also contained in K_s. However we know already from the previous argument that if α and β are separable over K, $K(\alpha, \beta)$ is separable (i.e. finite étale) over K. But then any element $\gamma \in K(\alpha, \beta)$ is separable over K. So in particular, $\alpha+\beta, \alpha\beta, \alpha^{-1}, \beta^{-1}, \ldots$, are all separable over K. Hence K_s is actually a subfield of $\bar{K}$.

This subfield K_s satisfies certain properties which we will later make into a definition.

(i) First, it is algebraic over K, since each element is the root of some polynomial in $K[x]$. Moreover, each element is separable over K. Both of these follow from the definition of K_s.

(ii) Second, K_s is separably closed, i.e. it has no finite separable extension except itself.

Proof. Let L be a finite separable extension of K_s. Let $\alpha \in L$ be an element of this extension; we would like to show that $\alpha \in K_s$. Consider the tower of extensions $K_s \subset K_s(\alpha) \subset L$. We know that α is separable over K_s (L is a separable extension of K_s), but we want that it is separable over K. Now consider the irreducible polynomial of α over K_s, $\mathrm{Irr}(\alpha, K_s) = \sum_{i=1}^{n} \beta_i x^i$. Since α is the root of a polynomial in $K(\beta_1, \beta_2, \ldots, \beta_n)[x]$, it is separable over this subfield $K(\beta_1, \beta_2, \ldots, \beta_n)$ of K_s. We now have a tower of extensions $K \subset K(\beta_1, \beta_2, \ldots, \beta_n) \subset K(\beta_1, \beta_2, \ldots, \beta_n, \alpha)$, each step of which is separable. Hence the whole thing is separable and hence α is separable over K.

Definition. *We call a field extension $\bar{L}$ of K a* **separable closure** *of K if it is algebraic over K, every element of $\bar{L}$ is separable over K, and $\bar{L}$ is separably closed.*

Our discussion above shows that every field has a separable closure, since we can just look at the subfield K_s of an algebraic closure. Moreover, a separable closure is unique in precisely the same fashion as an algebraic closure. Namely, if K, K' are fields with separable closures $\bar{K}, \bar{K}'$, respectively, and there exists $\sigma : K \to K'$ an isomorphism, then there exists an isomorphism $\tau \colon \bar{K} \to \bar{K}'$ of their separable closures which extends σ. As in the case of the algebraic closures, this extension τ is by no means unique, except in the case when it is not an extension at all—namely, when K, K' are already separably closed. In fact, if K_s and K_s' are separably closed, and $\bar{K}_s$, $\bar{K}_s'$ are algebraic closures, then τ is unique.

(A word about notation. From here on, we will place a bar over a field to designate either separable closure or algebraic closure. We hope that it will be clear from the context which is meant.)

11. Exercises

11.1. Let G be a finite abelian group, and let k be a field of characteristic not dividing the order of G. Denote by $k[G]$ the group ring of G over k, viewed as a k-algebra.

(a) For each $g \in G$, viewed as an element of $k[G]$, compute the trace $\mathrm{Tr}_{k[G]/k}(g)$, and prove that $k[G]$ is finite étale over k.

(b) Suppose that k is separably closed. Prove that $k[G]$ is, as a k-algebra, isomorphic to the product of $\#G$ copies of k, and deduce that the number of group homomorphisms $G \to k^*$ equals $\#G$.

11.2. Let R be a commutative non-zero ring, $R[t]$ the polynomial ring in one variable t over R, and M a free R-module of finite rank n. Let a be an R-endomorphism of M, and $f = \det(t - a) \in R[t]$ the *characteristic polynomial* of a; this is the determinant of the $R[t]$-endomorphism $1_M \otimes t - a \otimes 1$ $(= t - a)$ of the free $R[t]$-module $M \otimes_R R[t]$. Prove *Newton's identity*

$$\frac{f'}{f} = \sum_{m=0}^{\infty} \frac{\mathrm{Tr}(a^m)}{t^{m+1}}$$

in the ring $R[[1/t]][t]$ of formal Laurent series in $1/t$ over R. (*Hint.* One way to proceed is to rewrite this as $f'/f = \mathrm{Tr}\bigl(1/(t-a)\bigr)$, properly interpreted. To compute $\mathrm{Tr}\bigl(1/(t-a)\bigr)$, take a second variable u, and consider the coefficient of $\det\bigl(u - (t-a)^{-1}\bigr)$ at u^{n-1}.)

11.3. Let R be a commutative ring, and let A be the R-algebra $R[X]/(X^3 - X - 1)R[X]$. Prove that A is finite étale over R if and only if $23 \in R^*$.

11.4. Let R be a commutative ring, $f \in R[X]$ a monic polynomial, and A the R-algebra $R[X]/fR[X]$. Write $n = \deg f$ and $\alpha = (X \bmod f) \in A$.

(a) Prove that $1, \alpha, \ldots, \alpha^{n-1}$ is an R-basis for A, and that A is finite étale over R if and only if the determinant of the matrix $\bigl(\mathrm{Tr}_{A/R}(\alpha^{i+j})\bigr)_{i,j=0}^{n-1}$ belongs to R^*; that determinant is called the *discriminant* of f.

(b) Define $\pi\colon A \to R$ by $\pi\bigl(\sum_{i=0}^{n-1} r_i\alpha^i\bigr) = r_{n-1}$, for $r_0, r_1, \ldots, r_{n-1} \in R$. Prove that $\mathrm{Hom}_R(A, R)$ is A-free of rank 1, a basis being given by π.

(c) Prove that the characteristic polynomial of the R-linear map $A \to A$ sending x to αx equals $f(t)$.

11.5. Let R, f, A, n, α, π be as in Exercise 11.4.

(a) Let i be a non-negative integer, and denote by f' the derivative of f. Prove that the element $\sum_{m=i}^{\infty} \mathrm{Tr}_{A/R}(\alpha^m) \cdot t^{i-m-1} \cdot f(t)$ of $R[[1/t]][t]$ actually belongs to $R[t]$; that it has degree $< n$ in t; and that it equals the remainder of $f'(t) \cdot t^i$ upon division by $f(t)$. (*Hint.* Use Newton's identity from Exercise 11.2.)

(b) Prove: $\mathrm{Tr}_{A/R} = f'(\alpha) \cdot \pi$. Prove also that A is finite étale over R if and only if $f'(\alpha) \in A^*$, and if and only if $f'R[X] + fR[X] = R[X]$.

11.6. Let k be a field, p a prime number, and $a \in k^*$.

(a) Prove: $X^p - a$ is irreducible over k if and only if it has no zero in k.

(b) Prove: the trace map $k[X]/(X^p - a)k[X] \to k$ is zero if and only if $\operatorname{char} k = p$.

(c) Give an example of a finite inseparable (i. e., not separable) extension of fields.

11.7. Let R be a commutative ring that is *connected* (see Exercise 1.33).

(a) Let n be a non-negative integer, and let A be the totally split R-algebra R^n. Prove that any R-algebra homomorphism $A \to R$ is equal to one of the n projections.

(b) Prove that the category $\mathbf{Ts}_R$ of totally split R-algebras is anti-equivalent to the category **sets** of finite sets.

11.8. Let $\mathbf{R}$ be the field of real numbers, and let G be a group of order two.

(a) Prove that an $\mathbf{R}$-algebra A is finite étale if and only if there exist non-negative integers n, m such that $A \cong_{\mathbf{R}} \mathbf{R}^n \times \mathbf{C}^m$ (as $\mathbf{R}$-algebras).

(b) Prove that the category $\mathbf{FEt}_{\mathbf{R}}$ of finite étale $\mathbf{R}$-algebras is anti-equivalent to the category ${}_G\mathbf{sets}$ of finite G-sets, in which the morphisms are the G-maps.

11.9. Let R be a commutative ring, R' a commutative R-algebra, and let the notation M', f' be as in Exercise 10.14. Let A be an R-algebra that is finitely generated and projective as an R-module. In class we exhibited an isomorphism $\varphi\colon \operatorname{Hom}_R(A, R)' \xrightarrow{\sim} \operatorname{Hom}_{R'}(A', R')$.

(a) Verify that when the A-linear map $A \to \operatorname{Hom}_R(A, R)$ sending 1 to $\operatorname{Tr}_{A/R}$ is tensored with $1_{R'}$ and next composed with φ, we obtain the A'-linear map $A' \to \operatorname{Hom}_{R'}(A', R')$ sending 1 to $\operatorname{Tr}_{A'/R'}$. (You may take the result of Exercise 10.14 for granted.) Conclude: if A is finite étale over R, then A' is finite étale over R'.

(b) Suppose that R' is faithfully flat as an R-module (see Exercise 10.17). Prove: A is finite étale over R if and only if A' is finite étale over R'.

11.10. Let R be a commutative ring, and R' a commutative R-algebra that is faithfully flat over R. Let A be a commutative R-algebra, and write $A' = A \otimes_R R'$. Prove: A is finite étale over R if and only if A' is finite étale over R'. (You may take the result of Exercise 10.19(b) for granted.) (*Note.* This is a little sharper than Exercise 11.9(b), in which it is already assumed that A is finitely generated and projective as an R-module.)

11.11. (a) Supply the details of the following proof that every field k has an algebraic closure: the "infinite tensor product" $R = \bigotimes_f k[X]/fk[X]$ (taken over k), with f ranging over the non-constant polynomials $f \in k[X]$, is a non-zero k-algebra, and if $\mathfrak{m} \subset R$ is any maximal ideal, then $k' = R/\mathfrak{m}$ is an algebraic field extension of k in which every non-constant polynomial $f \in k[X]$ has a zero; the union of the fields $k \subset k' \subset k'' = (k')' \subset k''' \subset \ldots$ is now an algebraic closure of k.

(b) Supply the details of the following proof of the theorem that any field isomorphism $k \to l$ can be extended to an isomorphism between any algebraic closures K and L of k

and l, respectively: the ring $K \otimes_k L$ is non-zero, and for any maximal ideal $\mathfrak{n} \subset K \otimes_k L$ the natural maps $g: K \to (K \otimes_k L)/\mathfrak{n}$ and $h: L \to (K \otimes_k L)/\mathfrak{n}$ are isomorphisms; the required isomorphism $K \to L$ is now $h^{-1}g$.

11.12. A field k is called *perfect* if every finite extension $k \subset l$ is separable. Prove: k is perfect if and only if either $\operatorname{char} k = 0$, or $\operatorname{char} k = p > 0$ and the map $F: k \to k$ defined by $F(a) = a^p$ is a field automorphism.

(a) Give an example of a field that is not perfect.

(b) Let $k \subset l$ be a finite extension of fields. Prove: k is perfect if and only if l is perfect.

11.13. Let k be a field. An algebraic extension l of k (possibly infinite) is called *separable* if every element of l is separable over k. (In class we showed that, for *finite* extensions of fields, "separable" is the same as "finite étale".) A field extension K of k is called a *separable closure* of k if K is separably closed and separable algebraic over k.

(a) In class the existence of separable closures was proved. Prove: if k, l are fields, and K, L are separable closures of k, l, respectively, then any isomorphism $k \to l$ can be extended to an isomorphism $K \to L$.

(b) Let $k \subset l$ be a field extension, and let L be a separable closure of l. Prove that L contains a separable closure of k. Prove also that L itself is a separable closure of k if and only if l is separable algebraic extension of k.

12. The main theorem of Galois theory

We define a notion from the homework exercises which will show up later in our Galois theory.

We give the definition from Lang, Ch. I, Section 10. An *inversely directed family* of finite groups is a family $\{\pi_i\}_{i \in I}$ of finite groups π_i, where I is a directed partially ordered set, together with a family of group homomorphisms

$$f_i^j: \pi_j \to \pi_i,$$

one for each pair of elements i, $j \in I$ with $j \succeq i$, such that the following compatibility conditions hold: for $k \preceq i \preceq j$, we have

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

The inverse limit of the family is defined to be a closed subset of the product $\prod_{i \in I} \pi_i$, where each finite group π_i is given the discrete topology and the product is given the product topology. Namely, we let the inverse limit

$$\pi = \varprojlim_{i \in I} \pi_i$$

consist of elements $\langle x_i \rangle \in \prod_{i \in I} \pi_i$ such that for all i and $j \succeq i$ we have $f_i^j(x_j) = x_i$. This can be easily verified to be a closed subgroup of the product.

We call a group which is an inverse limit of finite groups a *profinite* group. Note that we can give π the subspace topology, which makes it into a topological group. It turns out that we can characterize, by topological conditions, those topological groups which can be written as an inverse limit of finite groups.

Theorem. *A topological group is profinite if and only if it is Hausdorff, compact, and totally disconnected.*

An example of a topological group which is compact and Hausdorff, but *not* totally disconnected and hence not profinite, is the circle group, $\mathbf{R}/\mathbf{Z}$. It is easy to see, at least, that any profinite topological group satisfies the three given conditions, but the other direction is harder. The interested reader can find further information in Hewitt and Ross, *Abstract Harmonic Analysis*.

Definition. *Let π be a profinite group which acts on a set S. We say this action is* **continuous** *if for all $s' \in S$ the set*

$$\{(\sigma, s) : \sigma s = s'\}$$

is open in the product topology on $\pi \times S$, where π has been given the profinite topology, and S is taken with the discrete topology.

In other words, the action is continuous if the map $\pi \times S \to S$ defining the action is continuous, if S is given the discrete topology and $\pi \times S$ is given the product topology.

Profinite groups are at the center of our discussion of Galois Theory. One profinite group in particular is at the core of our investigation.

Definition. *Let K be a field, and let $\bar{K}$ be a separable closure. The* **fundamental group** *or* **absolute Galois group** *of K is the group of K-automorphisms of $\bar{K}$.*

The fundamental group can be denoted alternatively as π, π_K, $\mathcal{G}$, or $\mathcal{G}_K$, depending on how obvious the field K in question is, and whether we are examining the group from a more topological or more algebraic perspective.

It is important to recognize that we have defined the fundamental group by using a not-necessarily-unique separable closure $\bar{K}$. It is reasonable to ask how a different choice of the separable closure would affect the fundamental group. We know that the separable closure of a field, while not unique, is unique up to (non-unique) isomorphism. It follows that the fundamental group is also unique up to an isomorphism that is itself unique up to an inner automorphism.

Examples.

(i) If $K = \mathbf{C}$, then $\bar{K} = \mathbf{C}$, so π is trivial.
(ii) If $K = \mathbf{R}$, and $\bar{K} = \mathbf{C}$, then the only $\mathbf{R}$-automorphisms of $\mathbf{C}$ are the identity and complex conjugation. So $\pi \cong \mathbf{Z}/2\mathbf{Z}$.

Let K be any field. We show that the fundamental group π of K can be viewed as a profinite group. Let F be the set of monic polynomials $f \in K[X]$ with $\gcd(f, f') = 1$, and for each $f \in F$ let $Z(f)$ be the set of zeroes of f in $\bar{K}$. For any polynomial $f \in F$, the set $Z(f)$ is finite, with cardinality equal to the degree of f, since each root of f lies in $\bar{K}$. Moreover, since each element of $\bar{K}$ is the root of a monic polynomial of this type, we have the equality

$$\bigcup_{f \in F} Z(f) = \bar{K}.$$

Furthermore, for any root α of f and any element $\sigma \in \pi$, the image $\sigma(\alpha)$ is also a root of f. Namely, if $f = \sum_i a_i x^i$, then we have

$$f(\sigma\alpha) = \sum a_i(\sigma(\alpha))^i = \sum \sigma(a_i)(\sigma(\alpha))^i = \sigma\big(\sum a_i \alpha^i\big) = \sigma(f(\alpha)) = \sigma(0) = 0.$$

Hence π acts on $Z(f)$ for each f. Let $G(f) \subset \operatorname{Sym} Z(f)$ be the image of the resulting map $\pi \to \operatorname{Sym} Z(f)$; the group $G(f)$ is called the *Galois group* of f. One now readily checks that the induced map from π to the projective limit of the groups $G(f)$ (with F ordered by divisibility) is an *isomorphism*. Since all groups $G(f)$ are finite, this shows that π has the structure of a profinite group. In particular, π has a natural *topology*, which is called the *Krull topology*, after the German algebraist Wolfgang Krull (1899–1971). In this topology on π, two elements σ and τ are "close" to each other if they agree on a "large" finite subset of $\bar{K}$. More formally, let E be a finite subset of $\bar{K}$, and let σ be an element of π. Then

$$\mathcal{U}_{\sigma,E} = \{\tau \in \pi : \tau|_E = \sigma\}$$

is open, and these $\mathcal{U}_{\sigma,E}$ form a basis for the topology on π. The reader should check that this is equivalent to the definition we gave earlier for the topology of a general profinite group.

We are now about to state and prove the Main Theorem of Galois theory. We should, however, ask the reader to remember that we have left many questions unanswered about the fundamental groups of fields. Indeed, we have yet to explore some basic concrete examples, such as the fundamental group of $\mathbf{Q}$.

Our version of the Main Theorem follows Grothendieck's formulation. Recall an earlier theorem that the category of totally split K-algebras is anti-equivalent to the category of finite sets. We now wish to generalize this statement. (In fact, the previous theorem will play an important role in the proof of the generalization.) First, we define the category

π-**FiniteSet** of finite sets equipped with a continuous action of the profinite group π, the automorphism group of the algebraic closure of K. The morphisms between objects in this category are the maps which preserve the π-action, i.e., given objects S, T in this category, the morphisms from S to T are the maps $f\colon S \to T$ satisfying $f(\sigma s) = \sigma f(s)$ for all $\sigma \in \pi$ and all $s \in S$. We can also define the category ${}_K\mathbf{Finite\acute{E}taleAlg}$ of finite étale K-algebras, with K-algebra homomorphisms as morphisms.

In the discussion of that theorem, we constructed functors F and G between the two categories in a natural way; namely, we took a totally split algebra to the set of K-algebra homomorphisms from A to K, and took a finite set to the K-algebra of maps from S to K. We now generalize this.

Denote by ${}_K\mathbf{Algebra}(A, \bar{K})$ the set of K-algebra homomorphisms from A to $\bar{K}$, and denote by $\pi{-}\mathbf{Set}(S, \bar{K})$ the K-algebra of maps from S to $\bar{K}$ which preserve the π-action. Consider the contravariant functor

$$\begin{aligned} F\colon {}_K\mathbf{Finite\acute{E}taleAlg} &\to \pi - \mathbf{FiniteSet} \\ A &\mapsto {}_K\mathbf{Algebra}(A, \bar{K}). \end{aligned}$$

Similarly, we consider the contravariant functor

$$\begin{aligned} G\colon \pi\text{-}\mathbf{FiniteSet} &\to {}_K\mathbf{Finite\acute{E}taleAlg} \\ S &\mapsto \pi - \mathbf{Set}(S, \bar{K}). \end{aligned}$$

For $F(A)$ to be an element of the category $\pi - \mathbf{FiniteSet}$, we need to define a π-action on ${}_K\mathbf{Algebra}(A, \bar{K})$. For $\sigma \in \pi$, and f a K-algebra homomorphism from A to $\bar{K}$, we let the action be composition; namely, we define

$$\sigma f = \sigma \circ f.$$

We will, of course, need to check that these functors are well-defined. We will first consider, however, the action of F and G on products and coproducts, since this will aid the proof by allowing us to reduce to our already-known cases.

We note first that for two finite étale K-algebras A_1 and A_2, we have

$$F(A_1 \times A_2) \cong F(A_1) \sqcup F(A_2).$$

This follows from the fact that the image of any element of $F(A_1 \times A_2)$ is a subfield of $\bar{K}$. Hence the result follows from our discussion of maximal ideals in product rings. Since we have already shown that any finite étale K-algebra A is (as a K-algebra) isomorphic to a product of finite separable field extensions of K, we can thus reduce our proof to the case where the K-algebra A *is* a finite separable field extension of K.

Similarly, we note that for two finite π-sets S_1 and S_2, each equipped with an action of π, we have

$$G(S_1 \sqcup S_2) \cong G(S_1) \times G(S_2).$$

Namely, given any morphism f from $S_1 \sqcup S_2$ to $\bar{K}$ which preserves the π-action, we can immediately define morphisms f_1 from S_1 to $\bar{K}$ which also preserves the π-action (simply by restricting to S_1), and similarly for f_2. An analogous construction holds for the other direction.

The reader might notice a symmetry between products and coproducts by the action of the functors F and G. In fact, this property must hold if F and G are to be contravariant equivalences, which we intend to prove in the Main Theorem.

We now show that F is well-defined.

Lemma. *The functor F, as defined above, is well-defined; i.e. given A a finite étale K-algebra, $F(A)$ is a finite set, and the π-action as defined above is continuous.*

Proof. Let A be an algebra in ${}_K\mathbf{Finite\acute{E}taleAlg}$. We define a $\bar{K}$-algebra, B, by

$$B = A_{\bar{K}} = A \otimes_K \bar{K}.$$

We have an isomorphism of sets

$$F(A) = {}_K\mathbf{Algebra}(A, \bar{K}) \cong {}_{\bar{K}}\mathbf{Algebra}(B, \bar{K})$$

by the natural map $f \mapsto f \otimes 1$. Since $\bar{K}$ is separably closed, we see that the set on the right is finite, and has cardinality $\dim_{\bar{K}}(B)$, which we know to be the same as $\dim_K(A)$. Therefore, $F(A)$ is finite with cardinality $\dim_K(A)$. This proves the first half of the lemma.

To show that the π-action is continuous, we need only show that the action is continuous in the special case when A is a finite separable field extension of K. This follows by our remarks on the action of F on products of algebras. Hence, without loss of generality, we let $A = L$, a finite separable field extension of K.

Let S be the set $F(L) = {}_K\mathbf{Algebra}(L, \bar{K})$. Then the cardinality of S is just the degree of L over K. Pick some $\lambda \in S$. Then λ is an injective K-algebra homomorphism from L to $\bar{K}$.

We claim that the π-action on S is transitive. Let μ be a second element of S. Since λ and μ are both injective, $\mu \circ \lambda^{-1}$ is a field isomorphism from λL to μL. We know that $\bar{K}$ is also a separable closure of λL and of μL. Thus, from the uniqueness of separable closures, we can extend $\mu \circ \lambda^{-1}$ to an automorphism $\sigma \in \pi$. (Notice that this extension clearly fixes K, and hence is in fact in π.) So by the π-action on S, we see that

$$\sigma\lambda = \mu,$$

and so π acts transitively on S.

Since this action is transitive, S is isomorphic (as a π-set) to π/ρ, where ρ is the stabilizer of λ. So in fact,

$$S \cong_\pi \pi/\rho.$$

An element $\sigma \in \pi$ is in the stabilizer of λ when $\sigma\lambda = \lambda$, i.e. for all elements $l \in L$ we have $\sigma(\lambda l) = \lambda l$. Hence σ fixes every element in λL pointwise. Thus, the stabilizer is the subgroup of automorphisms of $\bar{K}$ fixing the subfield λL, namely, $\rho = \mathrm{Aut}_{\lambda L}(\bar{K})$.

We claim that ρ is open in π. Namely, since λL is a finite extension of K, any element of ρ agrees with every other element on a finite set, namely, some basis set of λL viewed as a vector space over K. Moreover, since *every* element σ which fixes these basis elements must fix all of L, $\mathrm{Aut}_{\lambda L}(\bar{K})$ is a basis open set by definition of our topology, and thus ρ is open. We claim that the action of π on π/ρ is continuous. For any two cosets $\sigma_1\rho, \sigma_2\rho$, the set $\{\sigma \in \pi : \sigma\sigma_1\rho = \sigma_2\rho\} = \sigma_2\rho\sigma^{-1}$ is open (it is another basis open set, as can be checked). Since the inverse image under the continuous action is a union of these open sets, it is also open, and therefore the action is continuous.

The reader is asked to keep in mind the contructions in the above proof, since they will be used again in the first half of the proof of the Main Theorem. We will not complete the proof that G is well-defined before we prove the Main Theorem, since it will require a result which we prove en route to the results of the Main Theorem. We will, however, give an idea of the proof now, and reduce the proof of the well-definedness of G to this result. It will also serve to illustrate our claim that the previous results over separably closed fields will be crucial in our proof of the Main Theorem. We will be examining the following diagram:

$$\begin{array}{ccc} {}_K\textbf{FiniteÉtaleAlg} & \underset{G}{\overset{F}{\rightleftarrows}} & \pi - \textbf{FiniteSet} \\ {\scriptstyle H}\downarrow & & \downarrow \text{forgetful functor} \\ {}_{\bar{K}}\textbf{FiniteÉtaleAlg} & \underset{G}{\overset{F}{\rightleftarrows}} & \textbf{FiniteSet} \end{array}$$

where the functor H is our familiar extension of scalars functor taking A to $A_{\bar{K}} = A \otimes_K \bar{K}$, and the map from $\pi-$**FiniteSet** to **FiniteSet** is the (appropriate) forgetful functor.

Lemma. *If the above diagram commutes (using the G functors), then G is well-defined, i.e. given a π-set S, $G(S)$ is a finite étale K-algebra.*

Proof. Given S a π-set, notice that $G(S)_{\bar{K}} = HG(S)$, in the above diagram. We now go the other way around the diagram, first applying the forgetful functor and then taking the $\bar{K}$-algebra of maps from S to $\bar{K}$, yielding the object $\bar{K}^S$. Since the diagram commutes, $G(S)_{\bar{K}} = \bar{K}^S$, so certainly $G(S)_{\bar{K}}$ is finite étale (it is even totally split). But $G(S)$ is finite étale over K if and only if $G(S)_{\bar{K}}$ is finite étale over $\bar{K}$. Hence $G(S)$ is a finite étale K-algebra, as desired.

We will show that this diagram commutes (using the G functors) in the proof of the Main Theorem.

Before embarking on the proof, we present some examples.

Example. This is our "motivating" example, which we will use in our proof of the main theorem. Let K be itself separably closed. (For example, let K be the complex numbers $\mathbf{C}$.) The fundamental group of $\mathbf{C}$ is trivial, as noted earlier. As we have already observed, ${}_K\mathbf{Finite\acute{E}taleAlg}$ is just the category of totally split K-algebras, and this is anti-equivalent to the category **FiniteSet** of finite sets. Furthermore, it is clear that **FiniteSet** is equivalent to π-**FiniteSet** when π is trivial.

Our second example is somewhat more revealing, since in this case the fundamental group is non-trivial (though still quite simple!).

Example. Let K be the field of real numbers, $\mathbf{R}$, and let $\bar{K}$ be the field of complex numbers $\mathbf{C}$, such that the fundamental group π is isomorphic to $\mathbf{Z}/2\mathbf{Z}$. We can classify ${}_K\mathbf{Finite\acute{E}taleAlg}$ quite simply, since every finite étale algebra over $\mathbf{R}$ is isomorphic to

$$\underbrace{\mathbf{R} \times \mathbf{R} \times \cdots \times \mathbf{R}}_{a\times} \times \underbrace{\mathbf{C} \times \mathbf{C} \times \cdots \times \mathbf{C}}_{b\times} = \mathbf{R}^a \times \mathbf{C}^b$$

as an $\mathbf{R}$-algebra. This corresponds to the set of order $a + 2b$ consisting of a 1-cycles, and b 2-cycles, on which π acts in the obvious way.

There are two interesting things to notice in this last example. First, if A is a finite étale $\mathbf{R}$-algebra, then $\dim(A) = \#F(A)$ and the number of field constituents of A is the same as the number of π-orbits in $F(A)$. Both of these equalities will become more transparent as we explore the fundamental group further.

We now state the theorem.

Theorem. *Let K be a field, $\bar{K}$ a separable closure of K, and π the resulting fundamental group. Then the category ${}_K\mathbf{Finite\acute{E}taleAlg}$ of finite étale K-algebras is anti-equivalent to the category of π-* **FiniteSet** *of finite sets provided with a continuous action of π.*

Proof. We need to show that the composite functors FG and GF are isomorphic to the corresponding identity functors.

We first examine the object $GF(A)$: this is the set of maps from ${}_K\mathbf{Algebra}(A, \bar{K})$ to $\bar{K}$ which preserve the π-action defined on both sets. We note also that for *any* set S, $G(S)$ can naturally be given a K-algebra structure by pointwise addition and multiplication, and a homomorphism $f\colon K \to G(S)$ given by the constant functions. (The hard part, of course, is proving that it is finite étale.) In order to show that Φ_A is a K-algebra isomorphism, we will explicitly construct a map which will turn out to be an isomorphism.

We define $\Phi_A : A \to GF(A)$ by

$$\Phi_A(a) = (f \mapsto f(a)),$$

the evaluation homomorphism. We need to check, first, that each $\Phi_A(a)$ is indeed a map preserving the π-action. However, since the action of π on ${}_K\mathbf{Algebra}(A, \bar{K})$ is just composition, we have for any $f \in {}_K\mathbf{Algebra}(A, \bar{K}), a \in A, \sigma \in \pi$,

$$(\sigma f)(a) = (\sigma \circ f)(a) = \sigma(f(a)),$$

and so the π-action is preserved.

We claim that Φ_A is a K-algebra isomorphism for any finite étale K-algebra A. As in our proof that F is well-defined, we need only consider the case where $A = L$, a finite separable field extension of K, for the same reason. Hence we assume without loss of generality that $A = L$. Recall that π acts transitively on $S = F(L) \cong_\pi \pi/\rho$, where ρ is the stabilizer of some arbitrary element $\lambda \in S$.

We now define a new subfield $\bar{K}^\rho$ of $\bar{K}$,

$$\bar{K}^\rho = \{\alpha \in \bar{K} | \sigma\alpha = \alpha, \forall \sigma \in \rho\}.$$

The reader should check that this is indeed a subfield of $\bar{K}$. This is often called the "fixed field" of ρ. We claim that there exists a bijection between $\bar{K}^\rho$ and $GF(L)$. Recall that $GF(L) = \pi{-}\mathbf{FiniteSets}(\pi/\rho, \bar{K})$. Note that any map of π-sets $f : \pi/\rho \to \bar{K}$ is completely determined by where f sends the single element λ, since we have just shown that the action of π is transitive. Now notice that for any element $\sigma \in \rho$, we have

$$f(\sigma\lambda) = \sigma f(\lambda) = f(\lambda),$$

since the π-action is preserved, and since ρ is, by definition, the stabilizer of λ in π. We have just shown that the image of λ under *any* map $f \in GF(L)$ must lie in the fixed field $\bar{K}^\rho$. On the other hand, λ can be mapped to any such element in the fixed field $\bar{K}^\rho$. Hence there is a bijective correspondence between $GF(L)$ and $\bar{K}^\rho$. But it is straightforward to check that this is also a K-algebra homomorphism, where $\bar{K}^\rho$ is viewed as a K-algebra in the natural way.

We can now define a map $\Phi_L : L \to \bar{K}$, first by mapping any element l to the π-set map $f \mapsto f(l), f \in {}_K\mathbf{Algebra}(L, \bar{K})$, and then using the above bijective correspondence to identify the π-set map with the image of λ. As a map, this means that $\Phi_L(l) = \lambda(l)$. To prove that Φ_L is an isomorphism of K-algebras, it now suffices to show that $\bar{K}^\rho = \lambda L$.

It is clear that $\lambda L \subset \bar{K}^\rho$. Let α be an element in $\bar{K}^\rho$. We now apply a counting argument. We know that

$$\mathrm{Aut}_{\lambda L(\alpha)}(\bar{K}) = \mathrm{Aut}_{\lambda L}(\bar{K}),$$

by the definition of $\bar{K}^\rho$. If we look at the index of each of these in π, we see that

$$[\lambda L(\alpha) : K] = [L : K],$$

and so $[\lambda L(\alpha) : \lambda L] = 1$, which means that $\lambda L(\alpha) = \lambda L$. This shows that α is in fact in λL, so $\bar{K}^\rho \subset \lambda L$.

This shows that $K_S^\rho = \lambda L$, and therefore Φ_L is an isomorphism, and so $GF \cong \mathrm{id}_{{}_K\mathbf{FiniteÉtaleAlg}}$.

We have finished the first half of the proof of the Main Theorem; we now claim that the functor G is well-defined, and that $FG \cong \mathrm{id}_{\pi-\mathbf{FiniteSets}}$. More specifically, we claim that the map $S \to FG(S) = \mathrm{Alg}_K(G(S), \bar{K})$ given by sending each element s to the K-algebra homomorphism $(f \mapsto f(s))$ is an isomorphism.

Recall that we have already shown that it suffices to show that the diagram

$$\begin{tikzcd} {}_K\mathbf{FiniteÉtaleAlg} \arrow[r, shift left, "F"] \arrow[d, "H"'] & \pi - \mathbf{FiniteSet} \arrow[l, shift left, "G"] \arrow[d, "\text{forgetful functor}"] \\ {}_{\bar{K}}\mathbf{FiniteÉtaleAlg} \arrow[r, shift left, "F"] & \mathbf{FiniteSet} \arrow[l, shift left, "G"] \end{tikzcd}$$

commutes in order to show that G is well-defined. We now show that, in fact, this also suffices to show that $FG \cong \mathrm{id}$.

Let S be a π-set. We know from previous arguments that $S \cong \mathrm{Alg}_{\bar{K}}(\bar{K}^S, \bar{K})$, (where we have applied the forgetful functor to S). If the diagram commutes, then we have an isomorphism $S \cong \mathrm{Alg}_{\bar{K}}(G(S)_{\bar{K}}, \bar{K})$. We have already noted that $\mathrm{Alg}_K(G(S), \bar{K}) \cong \mathrm{Alg}_{\bar{K}}(G(S)_{\bar{K}}, \bar{K})$, as π-sets. But $\mathrm{Alg}_K(G(S), \bar{K}) = FG(S)$, by definition. Hence $S \cong FG(S)$. (We have not yet explicitly constructed the isomorphism which proves that the diagram commutes. Once we do so, one needs to check that the composition of the three isomorphisms mentioned above is indeed given by the natural map sending an element s of S to the K-algebra homomorphism $(f \mapsto f(s))$. This is straightforward.)

In order to show that this diagram commutes, we need for a π-set S an isomorphism

$$\pi - \mathbf{FiniteSet}(S, \bar{K}) \otimes_K \bar{K} \to \bar{K}^S.$$

It is actually straightforward to define a homomorphism Θ_S from the left-hand side to $\bar{K}^S$; $\pi-\mathbf{FiniteSet}(S, \bar{K})$ is a K-algebra of functions on S, and in particular can be seen as a subset of $\bar{K}^S$. Multiplying these elements by general elements in $\bar{K}$ defines a bilinear map and thence a K-algebra map from the tensor product. For a general set S, however, it is quite tricky to show that this Θ_S is an isomorphism. However, in the case where S actually "comes from" a K-algebra A, we can see that Θ_S is an isomorphism almost immediately.

We have already shown, in the proof that $GF \cong \mathrm{id}$, that the map Φ_A is an isomorphism. Hence we have the diagram

$$\begin{array}{ccc} \pi - \mathbf{FiniteSet}(S, \bar{K}) \otimes_K \bar{K} & \longrightarrow & \bar{K}^S \\ \Big\uparrow & & \updownarrow \\ A \otimes_K \bar{K} & = & B \end{array}$$

where B is just $A \otimes_K \bar{K}$ by definition, and the isomorphism between B and $\bar{K}^S$ comes from the fact that $\bar{K}$ is separably closed, and hence all finite étale $\bar{K}$-algebras are totally split. We will use this result to show that Θ_S is an isomorphism for a general set S.

Since S is a π-set, we can view it as a disjoint union of its orbits S_i under the π-action. Each S_i is a transitive π-set, and therefore is isomorphic (as a π-set) to π/ρ_i, where ρ_i is the stabilizer of some designated $x \in S_i$. In other words,

$$S \cong_\pi \coprod_i \pi/\rho_i$$

for some subgroups ρ_i of π. Since the action of π on S is continuous, it can be seen that the stabilizer of any element is actually *open*. So each ρ_i is open, and therefore contains some basis open set. Namely, $\mathrm{Aut}_{L_i}(\bar{K}) \subset \rho_i$ for some finite separable field extensions $K \subset L_i \subset \bar{K}$. We define a new π-set $T = \coprod_i \pi/\mathrm{Aut}_{L_i}(\bar{K})$. It is clear that $T \cong F(\prod_i L_i)$, and so by our earlier result Θ_T is an isomorphism.

Since $\rho_i \supset \mathrm{Aut}_{L_i}(\bar{K})$, we know that $\pi/\mathrm{Aut}_{L_i}(\bar{K})$ maps onto π/ρ_i, so T maps onto S. In particular we have a map $G(S) \hookrightarrow G(T)$, by composition with the projection map. We now characterize the set of elements of $G(T)$ which can be considered as images of elements of $G(S)$. Let $f \in G(T) = \pi - \mathbf{FiniteSets}(T, \bar{K})$. Then f is in the image of the inclusion map if and only if $f(t_1) = f(t_2)$, for all pairs t_1, t_2 with the same image in S (i.e. lie in the same coset $\rho_i/\mathrm{Aut}_{L_i}(\bar{K})$. In other words, we must have $(t_1, t_2) \in T \times_S T$, the fiber product of π-sets over S. Note that we also have inclusion maps g_1 and g_2 from π-$\mathbf{FiniteSet}(T, \bar{K})$ to π-$\mathbf{FiniteSet}(T \times_S T, \bar{K})$ by $g_1(f) = f \circ P_1$ and $g_2(f) = f \circ P_2$, where P_i is the projection to the ith coordinate. We now forget for a moment that the $G(S)$, $G(T)$, and $G(T \times_S T)$ have an algebra structure. We have an exact sequence of K-vector spaces

$$0 \longrightarrow G(S) \overset{\text{incl.}}{\hookrightarrow} G(T) \xrightarrow{g_1 - g_2} G(T \times_S T).$$

Notice that this is *just* a linear map of vector spaces; in particular, $g_1 - g_2$ maps 1 to 0. Tensoring everything with $\bar{K}$, and remembering that tensoring with a free module preserves exactness, we have the following diagram.

$$\begin{array}{ccccccc} 0 & \longrightarrow & G(S)_{\bar{K}} & \hookrightarrow & G(T)_{\bar{K}} & \hookrightarrow & G(T \times_S T)_{\bar{K}} \\ & & \Big\downarrow \Theta_S & & \Big\downarrow \Theta_T & & \Big\downarrow \Theta_{T \times_S T} \\ 0 & \longrightarrow & K^S & \hookrightarrow & K^T & \longrightarrow & K^{T \times_S T} \end{array}$$

By diagram-chasing, we see that to show Θ_S is an isomorphism, it would suffice to show that $\Theta_{T\times_S T}$ is an isomorphism. (We already know that Θ_T is an isomorphism.) To show that $\Theta_{T\times_S T}$ is an isomorphism, it is enough to show that the stabilizer of each (t_1, t_2) is of the form $\mathrm{Aut}_L(\bar{K}) \subset \pi$ for some finite separable field extension $K \subset L \subset \bar{K}$. If this were the case, then $T \times_S T = F(\prod L)$ for some set of L, and then $\Theta_{T\times_S T}$ will be an isomorphism for the same reasons as in the case of Θ_T. We already know that the stabilizers of both t_1 and t_2 in T are of this form. The stabilizer of (t_1, t_2) is their intersection, which is $\mathrm{Aut}_{LL'}(\bar{K})$. Since LL' is also a finite separable field extension, $\Theta_{T\times_S T}$ is an isomorphism.

Hence the diagram commutes, $FG \cong \mathrm{id}$, the categories are anti-equivalent, and the Main Theorem is proved.

Remark. Every open subgroup $\rho \subset \pi$ is of the form $\mathrm{Aut}_L(\bar{K})$. To see this, apply the functor G to the π-set, π/ρ, to see that

$$\pi/\rho \cong FG(\pi/\rho) \cong F(K^\rho) = \pi/\operatorname{Aut}_{\bar{K}^\rho}(\bar{K}).$$

We have thus shown that the category of finite étale K-algebras is anti-equivalent to the category of sets equipped with a π-action, where π is the fundamental group of K in its separable closure $\bar{K}$. The anti-equivalence comes from expressing a finite étale K-algebra A as a finite product

$$A = \prod_{i=1}^{t} L_i$$

(where each L_i is a finite separable extension of K), and by expressing a π-set S as a finite disjoint union

$$S = \coprod_{i=1}^{t} \pi/\rho_i$$

(where each ρ_i is an open subgroup of π). The anti-equivalence comes from identifying the L_i with the ρ_i by

$$L_i = \bar{K}^{\rho_i}, \quad \rho_i = \mathrm{Aut}_{L_i}(\bar{K}).$$

This gives us an inclusion-reversing bijection between finite separable extensions of K and open subgroups of π. (The inclusion-reversing nature of the bijection is a consequence of the anti-equivalence of the categories.) In fact there is a more general formulation of this bijection by extending it to a bijection between *all* separable extensions of K (not necessarily finite) and *closed* subgroups of π. This statement implies that, in fact, all open subgroups are *also* closed; this is a general fact about topological groups. Let ρ be an open subgroup. Then all cosets of ρ are open, and therefore the union of all the cosets (except ρ itself) is open, so ρ is closed.

This brings up some questions. Is the opposite true, that every closed subgroup is open? Certainly not, since the trivial subgroup is closed and yet certainly not open (since this would ascribe the discrete topology to π which is not the case unless π is finite). One might also ask whether there are subgroups which are neither open nor closed. The answer for profinite groups is in general yes. For example, suppose π is the direct product

$$\pi = (\mathbf{Z}/2\mathbf{Z})^{\infty} = \prod_{n=1}^{\infty} \mathbf{Z}/2\mathbf{Z}.$$

The subgroup $(\mathbf{Z}/2\mathbf{Z})^{(\infty)}$ is dense in π, and yet not equal to π. This shows that this subgroup must not be closed (and therefore also not open.) As another example, consider the profinite completion of the integers,

$$\pi = \hat{\mathbf{Z}} = \lim_{\leftarrow} \mathbf{Z}/n\mathbf{Z}$$

(where the integers are ordered by divisibility), and $\mathbf{Z}$ is imbedded in the natural way into $\hat{\mathbf{Z}}$, then $\mathbf{Z}$ forms a dense subgroup of $\hat{\mathbf{Z}}$ that is clearly not the whole group, showing that $\mathbf{Z}$ is not closed, and so also not open in $\hat{\mathbf{Z}}$.

The extended bijection above between the not-necessarily-finite separable extensions of K and closed subgroups of π comes from noticing that every separable extension is a union of finite separable extensions, and therefore corresponds to an intersection of open subgroups, which is also an intersection of closed subgroups, and therefore closed. One will notice that this correspondence implies that all closed subgroups of profinite groups are intersections of open subgroups. This can be seen by examining the closure of a subgroup.

Let π be a profinite group with components π_i, and let ρ be a subgroup. Then the closure $\bar{\rho}$ of ρ is the subgroup

$$\bar{\rho} = \lim_{\leftarrow} \rho_i$$

where ρ_i is the image of ρ under the projection map P_i from π to the i-th component π_i. This shows that $\bar{\rho}$ is the intersection of the open subgroups

$$\bar{\rho}_i = P_i^{-1}(\rho_i).$$

The proof of the extended bijection is straightforward and left as an exercise for the reader.

As another example of the use of our dictionary, we examine an arbitrary subgroup ρ of π. Let L be the fixed field of ρ, $L = \bar{K}^{\rho}$. Let $\mathrm{Aut}_L(\bar{K}) = \rho'$. Clearly, $\rho \subset \rho'$. We claim that $\rho' = \bar{\rho}$, the closure of ρ. Since ρ' is an open subgroup of π and hence also closed, we know that $\rho \subset \bar{\rho} \subset \rho'$. This immediately implies

$$\bar{K}^{\rho'} \subset \bar{K}^{\bar{\rho}} \subset \bar{K}^{\rho},$$

and since we defined ρ' so that $\bar{K}^{\rho} = \bar{K}^{\rho'}$, the inclusions are actually equalities, and hence $\bar{K}^{\rho'} = \bar{K}^{\bar{\rho}}$, and hence $\rho' = \bar{\rho}$.

In fact, suppose T is an arbitrary subset of π. Consider the fixed field $\bar{K}^T$ of S, and let $\rho' = \mathrm{Aut}_{\bar{K}^T}(\bar{K})$. By a similar argument, ρ' is seen to be the closure of the subgroup generated by T.

We can now travel easily between the two descriptions of the goings-on in the separably closed extension of K. However, we note that we do not know many concrete examples of this profinite group. Indeed, up until this point, we have in our arsenal only two examples: if K is already separably (or algebraically) closed, then π is the trivial group, and if $K = \mathbf{R}$, then $\pi = \mathbf{Z}/2\mathbf{Z}$. Both of these, being finite, illustrate little of the rich structure of the profinite group which has been our topic of discussion. It happens, however, that for a general field K it is extremely difficult to write down the structure of π. Nevertheless, in the case where K is a finite field, we can show that $\pi = \hat{\mathbf{Z}}$.

Let K be a finite field of size $q = p^n$. Our strategy for showing that $\pi = \hat{\mathbf{Z}}$ will be to examine carefully the role of a particular element in π, namely, the Frobenius map $\phi\colon \alpha \mapsto \alpha^q$. We first need to check that this is an element of π, i.e. that it is an automorphism of $\bar{K}$ fixing K. It is easy to see that ϕ fixes K, since 0 is automatically fixed, and K^* is cyclic of order $q-1$. Also, ϕ is a homomorphism of $\bar{K}$ since the p-th power map is always a field homomorphism in characteristic p, and ϕ is just a composition of p-th power maps. In fact, ϕ fixes exactly the elements of K and no others, since $x^q - x$ has at most q roots and $\#K = q$. Since any field homomorphism is injective, given any finite extension L of K, ϕ is also a surjection when restricted to L (since L is also finite). Hence ϕ is an automorphism.

We now look at the subgroup $\langle\phi\rangle$ of π generated by ϕ. From the observation above, $\bar{K}^{\langle\phi\rangle} = K$. But $\mathrm{Aut}_K(\bar{K}) = \pi$, so in fact $\bar{\langle\phi\rangle} = \pi$. In other words, topologically speaking, ϕ generates π.

We know that the closure of a cyclic subgroup of any profinite group is a homomorphic image of $\hat{\mathbf{Z}}$, by the map

$$\begin{array}{ccc} \hat{\mathbf{Z}} & \rightarrow & G \\ 1 & \mapsto & \phi. \end{array}$$

where ϕ is a generator of the dense subgroup of G. (That is, if G is procyclic, then $G \cong \hat{\mathbf{Z}}/a\hat{\mathbf{Z}}$ for some Steinitz number a.)

In the case of finite fields, we happen to know exactly the fixed field of ϕ^n for any integer $n \geq 0$. We have

$$\bar{K}^{\langle\phi^n\rangle} = \{\alpha \in \bar{K} | \alpha^{q^n} - \alpha = 0\},$$

and since the polynomial $x^{q^n} - x$ has no double roots (one can see this by calculating the derivative), we have that $\#\bar{K}^{\langle\phi^n\rangle} = q^n$, and hence $[\bar{K}^{\langle\phi^n\rangle} : K] = n$. From previous arguments, we know that $\bar{K}^{\bar{\langle\phi^n\rangle}} = \bar{K}^{\langle\phi^n\rangle}$, so $[\bar{K}^{\bar{\langle\phi^n\rangle}} : K] = n$ also. This gives us an induced isomorphism

$$\begin{array}{ccc} \hat{\mathbf{Z}} & \xrightarrow{f} & \pi \\ \downarrow & & \downarrow \\ \hat{\mathbf{Z}}/n\hat{\mathbf{Z}} & \longrightarrow & \pi/\langle\bar{\phi^n}\rangle, \end{array}$$

for each n. Since this is an isomorphism for each n, the map f is also an isomorphism, and hence $\pi \cong \hat{\mathbf{Z}}$.

Now we know the structure of the absolute Galois group of any finite field. We warn the reader again, however, that this explicit description of the absolute Galois group of a field is a rare phenomenon: for instance, no one can write down the absolute Galois group of the rationals $\mathbf{Q}$. It turns out, however, that in most practical instances, questions involving algebraic number fields (finite field extensions of $\mathbf{Q}$) can be answered if we know something about the finite "shadows" of π, i.e. the structure of the projections $G(f)$ for some polynomials f in $\mathbf{Q}[x]$.

We now take a look at the notion of a Galois extension – an idea heavily used in most presentations of classical Galois theory. From our standpoint, we see that the condition can be stated in many different ways. The generalization of the definition to the case of infinite extensions will be immediate and straightforward. We require a preliminary definition.

Definition. *Let L be a field extension of K. We say L is* **normal** *over K if for any $\alpha \in L$, the irreducible polynomial* $\mathrm{Irr}(\alpha, K)$ *splits linearly in L.*

Theorem. *Let L be a field extension of K. The following are equivalent.*

(i) L is finite over K, separable, and normal over K;

(ii) L is the splitting field of a polynomial f with $\gcd(f, f') = 1$*;*

(iii) L is an intermediate field of some separable closure of K corresponding to an open normal subgroup of π.

(iv) There exists a finite subgroup G of $\mathrm{Aut}(L)$ *such that $K = L^G$.*

(v) L is finite over K, and K is the fixed field of the relative automorphism group of L, i.e.

$$K = L^{\mathrm{Aut}_K(L)}.$$

Definition. *If all of the above (equivalent) conditions hold, then L over K is called* **finite Galois** *over K.*

To generalize the statement of the theorem to the case of (possibly) infinite extensions requires only a few changes. First, L is allowed to be an *algebraic* extension of K, not just a finite extension. In part (iii) of the theorem, L is allowed to correspond to a *closed* subgroups (i.e. an intersection of open subgroups) of π. In part (iv), the requirement is that there exists a *compact* subgroup G of $\mathrm{Aut}(L)$ such that the statement holds, where we give

Aut(L) the topology induced by the product topology on L^L, where each automorphism is viewed as a map of sets. (It is necessary to show that this is in fact a topological group.)

We can reformulate a classical version of the Main Theorem of Galois theory as follows. Let L be a Galois extension of a field K, and let $G = \mathrm{Gal}(L/K)$ be the corresponding topological Galois group. Then we have the following three correspondences:

$$\{\text{intermediate fields between } K \text{ and } L\} \leftrightarrow \{\text{closed subgroups of } G\}$$
$$\{\text{intermediate finite extensions}\} \leftrightarrow \{\text{open subgroups}\}$$
$$\{\text{intermediate Galois extensions}\} \leftrightarrow \{\text{normal subgroups}\}$$

Let L be a Galois extension of K, with Galois group G, and intermediate Galois extension M. The Galois groups are related by

$$K \overbrace{\underbrace{\subset}_{G/H} M \underbrace{\subset}_{H}}^{G} L$$

It is important to notice that this formulation has two "variables" as it were, both K and the Galois extension L. But in fact all information that can be found in this formulation is inferrable from the situation where L is $\bar{K}$, the separable closure of K. The Grothendieck formulation which we discussed therefore has the advantage that it eliminates the "unnecessary" second variable by setting it automatically to the most informative case. It has the additional advantage that it generalizes much more easily to the replacement of fields by commutative rings. One can track Galois theory through the years from being a discussion of polynomials, to an exploration of splitting fields, and finally to the Grothendieck formulation that we have used in this unit.

One well-known application of Galois theory is the study of solving polynomials by radicals. Let K be a field of characteristic 0, let f be a monic polynomial in $K[x]$, and let $Z(f)$ be the set of zeroes of the polynomial in the separable (algebraic) closure $\bar{K}$. We say that f is *solvable by radicals* if $K(Z(f))$ is a subset of K_t, which is the end of a finite sequence

$$K = K_0 \subset K_1 \subset \ldots \subset K_t$$

where $K_{i+1} = K_i(\alpha_i)$ and each $\alpha_i \in \bar{K}$ is such that $\alpha_i^{m_i} \in K_i$ for some positive integer m_i. By Galois theory we can determine that f is solvable by radicals if and only if $G(f)$ is solvable (as a group). If f has degree n, then $G(f)$ is a subgroup of S_n, the symmetric group on the roots of f, a set of n elements.

If f has degree n less than or equal to 4, since S_n (and any of its subgroups) is solvable, f is solvable by radicals. That these groups are solvable is an elementary group theory result. But equally elementary is the result that the alternating group A_5 is simple, so S_5 not solvable. Therefore a polynomial of degree greater than 4 is not in general solvable by radicals. (Certainly some are, such as $f(x) = x^5 - 2$, but others aren't. For example the polynomial $f(x) = x^n - x - 1$ has Galois group $G(f) = S_n$, and so is not solvable by radicals for any $n \geq 5$.)

Another famous application of Galois theory is toward the ancient Greek notion of constructibility. Let K be a subfield of the real numbers $\mathbf{R}$, and let f be a polynomial in $K[x]$. We say that f is *constructible* from K if f is solvable and each of the m_i from the "solution chain" is 2. Equivalently f is solvable if $G(f)$ has cardinality a power of 2. The reason why constructible polynomials are interesting is that they correspond to constructible objects in the complex plane using a ruler and compass. The constructible points in the complex plane are precisely the roots of the constructible polynomials.

One can see that for example the trisection of an angle corresponds to constructing the root of a cubic polynomial. Since a typical cubic polynomial f has a Galois group of order divisible by 3, most angle trisections are not constructible. (Some, such as the trisection of a right angle, are.) The Greeks asked what regular polygons are constructible. It is easy to see that a constructible polygon corresponds to constructing the roots of unity. Let f be the polynomial $f(x) = x^n - 1$. The factors of f are called *cyclotomic polynomials*, and $G(Z(f))$ is called a cyclotomic extension (over any field K such that $\operatorname{char} K$ does not divide n). We see that

$$G(f) \subset \operatorname{Aut}(\mathbf{Z}/n\mathbf{Z}) = \operatorname{End}(\mathbf{Z}/n\mathbf{Z})^* \cong (\mathbf{Z}/n\mathbf{Z})^*$$

with equality if $K = \mathbf{Q}$. The cardinality of $G(f)$ is the number of units in $\mathbf{Z}/n\mathbf{Z}$, which is the Euler function, $\varphi(n)$. So a regular n-gon is constructible if and only if $\varphi(n)$ is a power of 2. For example, since $\varphi(10) = 4$, a regular decagon is constructible. This the Greeks knew by actually constructing it. Since the Greeks did not know the theory behind these constructions, they couldn't go much further.

The applications and explorations of Galois theory go far beyond those enumerated in this unit. But it should provide a good launching point for those who are interested in a modern understanding of some of the basic results from Galois theory.

12. Exercises.

12.1. Let G be a group. The *profinite completion* $\hat{G}$ of G is defined to be the projective limit

$$\hat{G} = \lim_{\leftarrow} G/N,$$

with N ranging over the set of normal subgroups of finite index in G, ordered by $N \leq N'$ if and only if $N \supset N'$. This is a profinite group.

(a) Prove that there is a natural group homomorphism $f: G \to \hat{G}$, and that $f(G)$ is dense in $\hat{G}$. Can you give an example in which f is not injective?

(b) Let ${}_G\mathbf{sets}$ denote the category of finite sets provided with an action of G, and let ${}_{\hat{G}}\mathbf{sets}$ denote the category of finite sets provided with a continuous action of $\hat{G}$. Prove that these two categories are equivalent.

12.2. Let I be a partially ordered set that is *directed* in the sense that for any two elements i, $j \in I$ there exists $k \in I$ with $k \geq i$ and $k \geq j$. A system $\big((\pi_i)_{i\in I}, (f_i^j)_{i,j\in I, i\leq j}\big)$ of groups π_i and group homomorphisms $f_i^j: \pi_j \to \pi_i$ is said to be *directed* if one has $f_i^i = 1_{\pi_i}$ for every $i \in I$ and $f_i^j \circ f_j^k = f_i^k$ for any i, j, $k \in I$ with $i \leq j \leq k$.

Let $\big((\pi_i)_{i\in I}, (f_i^j)_{i,j\in I, i\leq j}\big)$ be a directed system in which every group π_i is *finite* and every f_i^j is *surjective*. Write $\pi = \lim_{\leftarrow} \pi_i$; this is a profinite group.

(a) Prove that for each $i \in I$ the natural map $\pi \to \pi_i$ is surjective.

(b) Suppose that π acts on a finite set S. Prove that the following are equivalent: (i) the action is continuous in the sense that the map $\pi \times S \to S$ defining the action is continuous, if S is given the discrete topology, π the profinite group topology, and $\pi \times S$ the product topology; (ii) for each $s \in S$ the stabilizer π_s is open in π; (iii) the kernel of the map $\pi \to \operatorname{Sym} S$ that defines the action is open; (iv) there exists $i \in I$ such that the action of π on S arises from an action of π_i on S through the map $\pi \to \pi_i$.

12.3. Let $\hat{\mathbf{Z}}$ be the profinite completion of the additive group of $\mathbf{Z}$.

(a) Prove that there is a unique ring structure on $\hat{\mathbf{Z}}$ for which (i) the ring operations are continuous maps $\hat{\mathbf{Z}} \times \hat{\mathbf{Z}} \to \hat{\mathbf{Z}}$, and (ii) the natural map $\mathbf{Z} \to \hat{\mathbf{Z}}$ is a ring homomorphism.

(b) Prove that $\hat{\mathbf{Z}}/n\hat{\mathbf{Z}} \cong \mathbf{Z}/n\mathbf{Z}$ for every positive integer n.

(c) Prove that every action of $\hat{\mathbf{Z}}$ on a finite set is continuous.

12.4. Let $\hat{\mathbf{Z}}$ be as in Exercise 12.3.

(a) Prove that $\hat{\mathbf{Z}}$ is flat as a $\mathbf{Z}$-module. Is it faithfully flat?

(b) What are the closed maximal ideals of $\hat{\mathbf{Z}}$? Does $\hat{\mathbf{Z}}$ have any maximal ideals that are not closed?

12.5. Let S be a finite set, $F(S)$ the free group determined by S (as in Chapter 5 of these notes), and $\hat{F}(S)$ its profinite completion (see Exercise 12.1). Prove that the natural map $F(S) \to \hat{F}(S)$ is injective.

12.6. Let k be a field, k_{s} a separable closure of k, and $\pi = \mathrm{Aut}_k k_{\mathrm{s}}$ the group of k-automorphisms of k_{s}. The purpose of part (a) of this exercise is to show that π has in a natural manner the structure of a profinite group. The resulting topology on π is often called the *Krull topology* (Wolfgang Krull, German algebraist, 1899–1971).

(a) Let F be the set of monic polynomials $f \in k[X]$ for which $\gcd(f, f') = 1$. For $f \in F$, the group π permutes the set $Z(f)$ of zeros of f in k_{s}, and we let the *Galois group* $G(f)$ of f be the image of the resulting map $\pi \to \mathrm{Sym}\, Z(f)$. Prove that the natural map $\pi \to \varprojlim G(f)$, with f ranging over F (ordered by divisibility), is a group isomorphism.

(b) Let $\rho \subset \pi$ be a subgroup. Prove that ρ is open in π if and only if there exists a finite extension l of k with $l \subset k_{\mathrm{s}}$ and $\mathrm{Aut}_l\, k_{\mathrm{s}} \subset \rho$ (deduce this directly from the definitions, and do not use the Main Theorem).

12.7. (a) Let $\mathbf{D}$ be the category whose objects are pairs (S, σ), where S is a finite set and σ belongs to $\mathrm{Sym}\, S$, a morphism $(S, \sigma) \to (T, \tau)$ being a map $f\colon S \to T$ with $f \circ \sigma = \tau \circ f$. Prove that the category $\mathbf{D}$ is equivalent to the category ${}_{\hat{\mathbf{Z}}}\mathbf{sets}$ of finite sets provided with a continuous action of $\hat{\mathbf{Z}}$.

(b) Let p be a prime number, and let $\mathbf{C}$ be the category of finite commutative rings R with $\sqrt{0}_R = \{0\}$ and $\#R$ equal to a power of p. Prove that $\mathbf{C}$ is anti-equivalent to $\mathbf{D}$.

12.8. Let $f\colon \pi \to \rho$ be a continuous bijective group homomorphism from a profinite group π to a profinite group ρ. Prove that f is an isomorphism of topological groups (i. e., that the inverse of f is also continuous).

12.9. Let p be a prime number. The *ring* $\mathbf{Z}_p$ *of p-adic integers* is defined to be the projective limit of the rings $\mathbf{Z}/p^n\mathbf{Z}$, with n ranging over the set of non-negative integers; the transition maps $\mathbf{Z}/p^m\mathbf{Z} \to \mathbf{Z}/p^n\mathbf{Z}$, for $m \geq n$, are the unique ring homomorphisms. Prove that $\mathbf{Z}_p$ is a local principal ideal domain, with maximal ideal $p\mathbf{Z}_p$, and with residue class field $\mathbf{F}_p$.

12.10. Prove that there is an isomorphism $\hat{\mathbf{Z}} \cong \prod_p \mathbf{Z}_p$ of topological rings (definition obvious), the product ranging over all prime numbers p, with $\mathbf{Z}_p$ as in Exercise 12.9.

12.11. A *Steinitz number* or *supernatural number* is a formal expression

$$a = \prod_p p^{a(p)},$$

where p ranges over the set of all prime numbers, and $a(p) \in \{0, 1, 2, \ldots, \infty\}$ for each p (Ernst Steinitz, German mathematician, 1871–1928). If $a = \prod_p p^{a(p)}$ is a Steinitz number, then we write $a\hat{\mathbf{Z}}$ for the subgroup of $\hat{\mathbf{Z}}$ that under the isomorphism $\hat{\mathbf{Z}} \cong \prod_p \mathbf{Z}_p$ from Exercise 12.10 corresponds to $\prod_p p^{a(p)}\mathbf{Z}_p$, with $p^\infty \mathbf{Z}_p = \{0\}$.

(a) Let a be a Steinitz number. Prove that $a\hat{\mathbf{Z}}$ is the intersection of all groups $n\hat{\mathbf{Z}}$, with n ranging over the positive integers that divide a (in an obvious sense).

(b) Prove that the map sending a to $a\hat{\mathbf{Z}}$ gives a bijection from the set of Steinitz numbers to the set of closed subgroups of $\hat{\mathbf{Z}}$. For which Steinitz numbers a is $a\hat{\mathbf{Z}}$ an *open* subgroup of $\hat{\mathbf{Z}}$?

(c) Let k be a finite field. Explain how Steinitz numbers classify k-isomorphism classes of algebraic field extensions l of k.

12.12. Let π be a profinite group. We call π *procyclic* if there exists $\sigma \in \pi$ such that the subgroup generated by σ is dense in π. Prove that the following assertions are equivalent: (i) π is procyclic; (ii) π is the inverse limit of a directed family of finite *cyclic* groups; (iii) $\pi \cong \hat{\mathbf{Z}}/a\hat{\mathbf{Z}}$ for some Steinitz number a (see Exercise 12.11); (iv) for any pair ρ, $\rho' \subset \pi$ of open subgroups of π with $\mathrm{index}(\pi : \rho) = \mathrm{index}(\pi : \rho')$ one has $\rho = \rho'$. Prove also that the Steinitz number a in (iii) is unique if it exists.

12.13. (a) Prove that every $a \in \hat{\mathbf{Z}}$ has a unique representation as an infinite sum $a = \sum_{n=1}^{\infty} c_n n!$, with $c_n \in \{0, 1, \ldots, n\}$.

(b) Let $b \in \mathbf{Z}$, $b \geq 0$, and define the sequence $(a_n)_{n=0}^{\infty}$ of non-negative integers by $a_0 = b$, $a_{n+1} = 2^{a_n}$. Prove that the limit $\lim_{n\to\infty} a_n$ exists in $\hat{\mathbf{Z}}$, and that it is independent of b.

(c) Let $a = \lim_{n\to\infty} a_n$ as in (b), and write $a = \sum_{n=1}^{\infty} c_n n!$ as in (a). Compute c_1, c_2, $\ldots$, c_{10}.

12.14. Let A be an abelian group with the property that every $a \in A$ has finite order (such a group is called a *torsion* group). Prove that A carries a unique $\hat{\mathbf{Z}}$-module structure.