Graduate Texts in Mathematics

Serge Lang

Algebra

Revised Third Edition

Volume 1

Springer

Graduate Texts in Mathematics 211

Books of Related Interest by Serge Lang

Math Talks for Undergraduates
1999, ISBN 0-387-98749-5

Linear Algebra, Third Edition
1987, ISBN 0-387-96412-6

Undergraduate Algebra, Second Edition
1990, ISBN 0-387-97279-X

Undergraduate Analysis, Second Edition
1997, ISBN 0-387-94841-4

Complex Analysis, Third Edition
1993, ISBN 0-387-97886

Real and Functional Analysis, Third Edition
1993, ISBN 0-387-94001-4

Algebraic Number Theory, Second Edition
1994, ISBN 0-387-94225-4

Other Books by Lang Published by Springer-Verlag

Introduction to Arakelov Theory • Riemann-Roch Algebra (with William Fulton) • Complex Multiplication • Introduction to Modular Forms • Modular Units (with Daniel Kubert) • Fundamentals of Diophantine Geometry • Elliptic Functions • Number Theory III • Survey of Diophantine Geometry • Fundamentals of Differential Geometry • Cyclotomic Fields I and II • $SL_2(\mathbf{R})$ • Abelian Varieties • Introduction to Algebraic and Abelian Functions • Introduction to Diophantine Approximations • Elliptic Curves: Diophantine Analysis • Introduction to Linear Algebra • Calculus of Several Variables • First Course in Calculus • Basic Mathematics • Geometry: A High School Course (with Gene Murrow) • Math! Encounters with High School Students • The Beauty of Doing Mathematics • THE FILE • CHALLENGES

Serge Lang

Algebra

Revised Third Edition

Springer

Serge Lang
Department of Mathematics
Yale University
New Haven, CT 96520
USA

Mathematics Subject Classification (2000): 13-01, 15-01, 16-01, 20-01

Library of Congress Cataloging-in-Publication Data
Algebra / Serge Lang.—Rev. 3rd ed.
p. cm.—(Graduate texts in mathematics; 211)
Includes bibliographical references and index.
ISBN 978-1-4612-6551-1 ISBN 978-1-4613-0041-0 (eBook)
DOI 10.1007/978-1-4613-0041-0
1. Algebra. I. Title. II. Series.
QA154.3.L3 2002
512—dc21 2001054916

ISBN 978-1-4612-6551-1

Printed on acid-free paper.

This title was previously published by Addison-Wesley, Reading, MA 1993.

9 8 (corrected printing 2005)

springer.com

FOREWORD

The present book is meant as a basic text for a one-year course in algebra, at the graduate level.

A perspective on algebra

As I see it, the graduate course in algebra must primarily prepare students to handle the algebra which they will meet in all of mathematics: topology, partial differential equations, differential geometry, algebraic geometry, analysis, and representation theory, not to speak of algebra itself and algebraic number theory with all its ramifications. Hence I have inserted throughout references to papers and books which have appeared during the last decades, to indicate some of the directions in which the algebraic foundations provided by this book are used; I have accompanied these references with some motivating comments, to explain how the topics of the present book fit into the mathematics that is to come subsequently in various fields; and I have also mentioned some unsolved problems of mathematics in algebra and number theory. The *abc* conjecture is perhaps the most spectacular of these.

Often when such comments and examples occur out of the logical order, especially with examples from other branches of mathematics, of necessity some terms may not be defined, or may be defined only later in the book. I have tried to help the reader not only by making cross-references within the book, but also by referring to other books or papers which I mention explicitly.

I have also added a number of exercises. On the whole, I have tried to make the exercises complement the examples, and to give them aesthetic appeal. I have tried to use the exercises also to drive readers toward variations and applications of the main text, as well as toward working out special cases, and as openings toward applications beyond this book.

Organization

Unfortunately, a book must be projected in a totally ordered way on the page axis, but that's not the way mathematics "is", so readers have to make choices how to reset certain topics in parallel for themselves, rather than in succession.

I have inserted cross-references to help them do this, but different people will make different choices at different times depending on different circumstances.

The book splits naturally into several parts. The first part introduces the basic notions of algebra. After these basic notions, the book splits in two major directions: the direction of algebraic equations including the Galois theory in Part II; and the direction of linear and multilinear algebra in Parts III and IV. There is some sporadic feedback between them, but their unification takes place at the next level of mathematics, which is suggested, for instance, in §15 of Chapter VI. Indeed, the study of algebraic extensions of the rationals can be carried out from two points of view which are complementary and interrelated: representing the Galois group of the algebraic closure in groups of matrices (the linear approach), and giving an explicit determination of the irrationalities generating algebraic extensions (the equations approach). At the moment, representations in GL_2 are at the center of attention from various quarters, and readers will see GL_2 appear several times throughout the book. For instance, I have found it appropriate to add a section describing all irreducible characters of $GL_2(F)$ when F is a finite field. Ultimately, GL_2 will appear as the simplest but typical case of groups of Lie types, occurring both in a differential context and over finite fields or more general arithmetic rings for arithmetic applications.

After almost a decade since the second edition, I find that the basic topics of algebra have become stable, with one exception. I have added two sections on elimination theory, complementing the existing section on the resultant. Algebraic geometry having progressed in many ways, it is now sometimes returning to older and harder problems, such as searching for the effective construction of polynomials vanishing on certain algebraic sets, and the older elimination procedures of last century serve as an introduction to those problems.

Except for this addition, the main topics of the book are unchanged from the second edition, but I have tried to improve the book in several ways.

First, some topics have been reordered. I was informed by readers and reviewers of the tension existing between having a textbook usable for relatively inexperienced students, and a reference book where results could easily be found in a systematic arrangement. I have tried to reduce this tension by moving all the homological algebra to a fourth part, and by integrating the commutative algebra with the chapter on algebraic sets and elimination theory, thus giving an introduction to different points of view leading toward algebraic geometry.

The book as a text and a reference

In teaching the course, one might wish to push into the study of algebraic equations through Part II, or one may choose to go first into the linear algebra of Parts III and IV. One semester could be devoted to each, for instance. The chapters have been so written as to allow maximal flexibility in this respect, and I have frequently committed the crime of lèse-Bourbaki by repeating short arguments or definitions to make certain sections or chapters logically independent of each other.

Granting the material which under no circumstances can be omitted from a basic course, there exist several options for leading the course in various directions. It is impossible to treat all of them with the same degree of thoroughness. The precise point at which one is willing to stop in any given direction will depend on time, place, and mood. However, any book with the aims of the present one must include a choice of topics, pushing ahead·in deeper waters, while stopping short of full involvement.

There can be no universal agreement on these matters, not even between the author and himself. Thus the concrete decisions as to what to include and what not to include are finally taken on grounds of general coherence and aesthetic balance. Anyone teaching the course will want to impress their own personality on the material, and may push certain topics with more vigor than I have, at the expense of others. Nothing in the present book is meant to inhibit this.

Unfortunately, the goal to present a fairly comprehensive perspective on algebra required a substantial increase in size from the first to the second edition, and a moderate increase in this third edition. These increases require some decisions as to what to omit in a given course.

Many shortcuts can be taken in the presentation of the topics, which admits many variations. For instance, one can proceed into field theory and Galois theory immediately after giving the basic definitions for groups, rings, fields, polynomials in one variable, and vector spaces. Since the Galois theory gives very quickly an impression of depth, this is very satisfactory in many respects.

It is appropriate here to recall my original indebtedness to Artin, who first taught me algebra. The treatment of the basics of Galois theory is much influenced by the presentation in his own monograph.

Audience and background

As I already stated in the forewords of previous editions, the present book is meant for the graduate level, and I expect most of those coming to it to have had suitable exposure to some algebra in an undergraduate course, or to have appropriate mathematical maturity. I expect students taking a graduate course to have had some exposure to vector spaces, linear maps, matrices, and they will no doubt have seen polynomials at the very least in calculus courses.

My books *Undergraduate Algebra* and *Linear Algebra* provide more than enough background for a graduate course. Such elementary texts bring out in parallel the two basic aspects of algebra, and are organized differently from the present book, where both aspects are deepened. Of course, some aspects of the linear algebra in Part III of the present book are more "elementary" than some aspects of Part II, which deals with Galois theory and the theory of polynomial equations in several variables. Because Part II has gone deeper into the study of algebraic equations, of necessity the parallel linear algebra occurs only later in the total ordering of the book. Readers should view both parts as running simultaneously.

Unfortunately, the amount of algebra which one should ideally absorb during this first year in order to have a proper background (irrespective of the subject in which one eventually specializes) exceeds the amount which can be covered physically by a lecturer during a one-year course. Hence more material must be included than can actually be handled in class. I find it essential to bring this material to the attention of graduate students.

I hope that the various additions and changes make the book easier to use as a text. By these additions, I have tried to expand the general mathematical perspective of the reader, insofar as algebra relates to other parts of mathematics.

Acknowledgements

I am indebted to many people who have contributed comments and criticisms for the previous editions, but especially to Daniel Bump, Steven Krantz, and Diane Meuser, who provided extensive comments as editorial reviewers for Addison-Wesley. I found their comments very stimulating and valuable in preparing this third edition. I am much indebted to Barbara Holland for obtaining these reviews when she was editor. I am also indebted to Karl Matsumoto who supervised production under very trying circumstances. I thank the many people who have made suggestions and corrections, especially George Bergman and students in his class, Chee-Whye Chin, Ki-Bong Nam, David Wasserman, Randy Scott, Thomas Shiple, Paul Vojta, Bjorn Poonen and his class, in particular Michael Manapat.

For the 2002 and beyond Springer printings

From now on, *Algebra* appears with Springer-Verlag, like the rest of my books. With this change, I considered the possibility of a new edition, but decided against it. I view the book as very stable. The only addition which I would make, if starting from scratch, would be some of the algebraic properties of SL_n and GL_n (over $\mathbf{R}$ or $\mathbf{C}$), beyond the proof of simplicity in Chapter XIII. As things stood, I just inserted some exercises concerning some aspects which everybody should know. The material actually is now inserted in a new edition of *Undergraduate Algebra*, where it properly belongs. The algebra appears as a supporting tool for doing analysis on Lie groups, cf. for instance Jorgenson/ Lang *Spherical Inversion on* $SL_n(\mathbf{R})$, Springer Verlag 2001.

I thank specifically Tom von Foerster, Ina Lindemann and Mark Spencer for their editorial support at Springer, as well as Terry Kornak and Brian Howe who have taken care of production.

Serge Lang
New Haven 2004

Logical Prerequisites

We assume that the reader is familiar with sets, and with the symbols $\cap$, $\cup$, $\supset$, $\subset$, $\in$. If A, B are sets, we use the symbol $A \subset B$ to mean that A is contained in B but may be equal to B. Similarly for $A \supset B$.

If $f: A \to B$ is a mapping of one set into another, we write

$$x \mapsto f(x)$$

to denote the effect of f on an element x of A. We distinguish between the arrows $\to$ and $\mapsto$. We denote by $f(A)$ the set of all elements $f(x)$, with $x \in A$.

Let $f: A \to B$ be a mapping (also called a map). We say that f is **injective** if $x \neq y$ implies $f(x) \neq f(y)$. We say f is **surjective** if given $b \in B$ there exists $a \in A$ such that $f(a) = b$. We say that f is **bijective** if it is both surjective and injective.

A subset A of a set B is said to be **proper** if $A \neq B$.

Let $f: A \to B$ be a map, and A' a subset of A. The restriction of f to A' is a map of A' into B denoted by $f \,|\, A'$.

If $f: A \to B$ and $g: B \to C$ are maps, then we have a composite map $g \circ f$ such that $(g \circ f)(x) = g(f(x))$ for all $x \in A$.

Let $f: A \to B$ be a map, and B' a subset of B. By $f^{-1}(B')$ we mean the subset of A consisting of all $x \in A$ such that $f(x) \in B'$. We call it the **inverse image** of B'. We call $f(A)$ the **image** of f.

A **diagram**

$$\begin{tikzcd} A \arrow[rr, "f"] \arrow[dr, "h"'] & & B \arrow[dl, "g"] \\ & C & \end{tikzcd}$$

is said to be **commutative** if $g \circ f = h$. Similarly, a **diagram**

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "\varphi"'] & B \arrow[d, "g"] \\ C \arrow[r, "\psi"'] & D \end{tikzcd}$$

is said to be **commutative** if $g \circ f = \psi \circ \varphi$. We deal sometimes with more complicated diagrams, consisting of arrows between various objects. Such diagrams are called commutative if, whenever it is possible to go from one object to another by means of two sequences of arrows, say

$$A_1 \xrightarrow{f_1} A_2 \xrightarrow{f_2} \cdots \xrightarrow{f_{n-1}} A_n$$

and

$$A_1 \xrightarrow{g_1} B_2 \xrightarrow{g_2} \cdots \xrightarrow{g_{m-1}} B_m = A_n,$$

then

$$f_{n-1} \circ \cdots \circ f_1 = g_{m-1} \circ \cdots \circ g_1,$$

in other words, the composite maps are equal. Most of our diagrams are composed of triangles or squares as above, and to verify that a diagram consisting of triangles or squares is commutative, it suffices to verify that each triangle and square in it is commutative.

We assume that the reader is acquainted with the integers and rational numbers, denoted respectively by **Z** and **Q**. For many of our examples, we also assume that the reader knows the real and complex numbers, denoted by **R** and **C**.

Let A and I be two sets. By a family of elements of A, indexed by I, one means a map $f: I \to A$. Thus for each $i \in I$ we are given an element $f(i) \in A$. Although a family does not differ from a map, we think of it as determining a collection of objects from A, and write it often as

$$\{f(i)\}_{i \in I}$$

or

$$\{a_i\}_{i \in I},$$

writing a_i instead of $f(i)$. We call I the indexing set.

We assume that the reader knows what an equivalence relation is. Let A be a set with an equivalence relation, let E be an equivalence class of elements of A. We sometimes try to define a map of the equivalence classes into some set B. To define such a map f on the class E, we sometimes first give its value on an element $x \in E$ (called a representative of E), and then show that it is independent of the choice of representative $x \in E$. In that case we say that f is **well defined**.

We have products of sets, say finite products $A \times B$, or $A_1 \times \cdots \times A_n$, and products of families of sets.

We shall use Zorn's lemma, which we describe in Appendix 2.

We let $\#(S)$ denote the number of elements of a set S, also called the **cardinality** of S. The notation is usually employed when S is finite. We also write $\#(S) = \text{card}(S)$.

CONTENTS

Part One The Basic Objects of Algebra

Part Two Algebraic Equations

Part One

THE BASIC OBJECTS OF ALGEBRA

This part introduces the basic notions of algebra, and the main difficulty for the beginner is to absorb a reasonable vocabulary in a short time. None of the concepts is difficult, but there is an accumulation of new concepts which may sometimes seem heavy.

To understand the next parts of the book, the reader needs to know essentially only the basic definitions of this first part. Of course, a theorem may be used later for some specific and isolated applications, but on the whole, we have avoided making long logical chains of interdependence.

CHAPTER I

Groups

§1. MONOIDS

Let S be a set. A mapping

$$S \times S \to S$$

is sometimes called a **law of composition** (of S into itself). If x, y are elements of S, the image of the pair (x, y) under this mapping is also called their **product** under the law of composition, and will be denoted by xy. (Sometimes, we also write $x \cdot y$, and in many cases it is also convenient to use an additive notation, and thus to write $x + y$. In that case, we call this element the **sum** of x and y. It is customary to use the notation $x + y$ only when the relation $x + y = y + x$ holds.)

Let S be a set with a law of composition. If x, y, z are elements of S, then we may form their product in two ways: $(xy)z$ and $x(yz)$. If $(xy)z = x(yz)$ for all x, y, z in S then we say that the law of composition is **associative**.

An element e of S such that $ex = x = xe$ for all $x \in S$ is called a **unit element**. (When the law of composition is written additively, the unit element is denoted by 0, and is called a **zero element**.) A unit element is unique, for if e' is another unit element, we have

$$e = ee' = e'$$

by assumption. In most cases, the unit element is written simply 1 (instead of e). For most of this chapter, however, we shall write e so as to avoid confusion in proving the most basic properties.

A **monoid** is a set G, with a law of composition which is associative, and having a unit element (so that in particular, G is not empty).

S. Lang, *Algebra*

Let G be a monoid, and $x_1, \ldots, x_n$ elements of G (where n is an integer > 1). We define their product inductively:

$$\prod_{\nu=1}^{n} x_\nu = x_1 \cdots x_n = (x_1 \cdots x_{n-1})x_n.$$

We then have the following rule:

$$\prod_{\mu=1}^{m} x_\mu \cdot \prod_{\nu=1}^{n} x_{m+\nu} = \prod_{\nu=1}^{m+n} x_\nu,$$

which essentially asserts that we can insert parentheses in any manner in our product without changing its value. The proof is easy by induction, and we shall leave it as an exercise.

One also writes

$$\prod_{m+1}^{m+n} x_\nu \quad \text{instead of} \quad \prod_{\nu=1}^{n} x_{m+\nu}$$

and we define

$$\prod_{\nu=1}^{0} x_\nu = e.$$

As a matter of convention, we agree also that the empty product is equal to the unit element.

It would be possible to define more general laws of composition, i.e. maps $S_1 \times S_2 \to S_3$ using arbitrary sets. One can then express associativity and commutativity in any setting for which they make sense. For instance, for commutativity we need a law of composition

$$f: S \times S \to T$$

where the two sets of departure are the same. **Commutativity** then means $f(x, y) = f(y, x)$, or $xy = yx$ if we omit the mapping f from the notation. For associativity, we leave it to the reader to formulate the most general combination of sets under which it will work. We shall meet special cases later, for instance arising from maps

$$S \times S \to S \quad \text{and} \quad S \times T \to T.$$

Then a product $(xy)z$ makes sense with $x \in S$, $y \in S$, and $z \in T$. The product $x(yz)$ also makes sense for such elements x, y, z and thus it makes sense to say that our law of composition is associative, namely to say that for all x, y, z as above we have $(xy)z = x(yz)$.

If the law of composition of G is commutative, we also say that G is **commutative (or abelian)**.

Let G be a commutative monoid, and $x_1, \dots, x_n$ elements of G. Let ψ be a bijection of the set of integers $(1, \dots, n)$ onto itself. Then

$$\prod_{\nu=1}^{n} x_{\psi(\nu)} = \prod_{\nu=1}^{n} x_\nu.$$

We prove this by induction, it being obvious for $n = 1$. We assume it for $n - 1$. Let k be an integer such that $\psi(k) = n$. Then

$$\begin{aligned}\prod_{1}^{n} x_{\psi(\nu)} &= \prod_{1}^{k-1} x_{\psi(\nu)} \cdot x_{\psi(k)} \cdot \prod_{1}^{n-k} x_{\psi(k+\nu)} \\ &= \prod_{1}^{k-1} x_{\psi(\nu)} \cdot \prod_{1}^{n-k} x_{\psi(k+\nu)} \cdot x_{\psi(k)}.\end{aligned}$$

Define a map φ of $(1, \dots, n - 1)$ into itself by the rule

$$\begin{aligned}\varphi(\nu) &= \psi(\nu) &&\text{if } \nu < k, \\ \varphi(\nu) &= \psi(\nu + 1) &&\text{if } \nu \geqq k.\end{aligned}$$

Then

$$\begin{aligned}\prod_{1}^{n} x_{\psi(\nu)} &= \prod_{1}^{k-1} x_{\varphi(\nu)} \prod_{1}^{n-k} x_{\varphi(k-1+\nu)} \cdot x_n \\ &= \prod_{1}^{n-1} x_{\varphi(\nu)} \cdot x_n,\end{aligned}$$

which, by induction, is equal to $x_1 \cdots x_n$, as desired.

Let G be a commutative monoid, let I be a set, and let $f : I \to G$ be a mapping such that $f(i) = e$ for almost all $i \in I$. (Here and thereafter, **almost all** will mean *all but a finite number*.) Let I_0 be the subset of I consisting of those i such that $f(i) \neq e$. By

$$\prod_{i \in I} f(i)$$

we shall mean the product

$$\prod_{i \in I_0} f(i)$$

taken in any order (the value does not depend on the order, according to the preceding remark). It is understood that the empty product is equal to e.

When G is written additively, then instead of a product sign, we write the sum sign Σ.

There are a number of formal rules for dealing with products which it would be tedious to list completely. We give one example. Let I, J be two sets, and

$f: I \times J \to G$ a mapping into a commutative monoid which takes the value e for almost all pairs (i, j). Then

$$\prod_{i \in I}\left[\prod_{j \in J} f(i, j)\right] = \prod_{j \in J}\left[\prod_{i \in I} f(i, j)\right].$$

We leave the proof as an exercise.

As a matter of notation, we sometimes write $\prod f(i)$, omitting the signs $i \in I$, if the reference to the indexing set is clear.

Let x be an element of a monoid G. For every integer $n \geqq 0$ we define x^n to be

$$\prod_1^n x,$$

so that in particular we have $x^0 = e$, $x^1 = x$, $x^2 = xx, \ldots$. We obviously have $x^{(n+m)} = x^n x^m$ and $(x^n)^m = x^{nm}$. Furthermore, from our preceding rules of associativity and commutativity, if x, y are elements of G such that $xy = yx$, then $(xy)^n = x^n y^n$. We leave the formal proof as an exercise.

If S, S' are two subsets of a monoid G, then we define SS' to be the subset consisting of all elements xy, with $x \in S$ and $y \in S'$. Inductively, we can define the product of a finite number of subsets, and we have associativity. For instance, if S, S', S'' are subsets of G, then $(SS')S'' = S(S'S'')$. Observe that $GG = G$ (because G has a unit element). If $x \in G$, then we define xS to be $\{x\}S$, where $\{x\}$ is the set consisting of the single element x. Thus xS consists of all elements xy, with $y \in S$.

By a **submonoid** of G, we shall mean a subset H of G containing the unit element e, and such that, if $x, y \in H$ then $xy \in H$ (we say that H is **closed** under the law of composition). It is then clear that H is itself a monoid, under the law of composition induced by that of G.

If x is an element of a monoid G, then the subset of powers x^n ($n = 0, 1, \ldots$) is a submonoid of G.

The set of integers $\geqq 0$ under addition is a monoid.

Later we shall define rings. If R is a commutative ring, we shall deal with multiplicative subsets S, that is subsets containing the unit element, and such that if $x, y \in S$ then $xy \in S$. Such subsets are monoids.

A routine example. Let **N** be the natural numbers, i.e. the integers $\geqq 0$. Then **N** is an additive monoid. In some applications, it is useful to deal with a multiplicative version. See the definition of polynomials in Chapter II, §3, where a higher-dimensional version is also used for polynomials in several variables.

An interesting example. We assume that the reader is familiar with the terminology of elementary topology. Let M be the set of homeomorphism classes of compact (connected) surfaces. We shall define an addition in M. Let S, S' be compact surfaces. Let D be a small disc in S, and D' a small disc in S'. Let C, C' be the circles which form the boundaries of D and D' respectively. Let D_0, D'_0 be the interiors of D and D' respectively, and glue $S - D_0$ to $S' - D'_0$ by identifying C with C'. It can be shown that the resulting surface is independent, up to homeomorphism, of the various choices made in the preceding construction. If σ, σ' denote the homeomorphism classes of S and S' respectively, we define $\sigma + \sigma'$ to be the class of the surface obtained by the preceding gluing process. It can be shown that this addition defines a monoid structure on M, whose unit element is the class of the ordinary 2-sphere. Furthermore, if τ denotes the class of the torus, and π denotes the class of the projective plane, then every element σ of M has a unique expression of the form

$$\sigma = n\tau + m\pi$$

where n is an integer $\geqq 0$ and $m = 0$, 1, or 2. We have $3\pi = \tau + \pi$.

(The reasons for inserting the preceding example are twofold: First to relieve the essential dullness of the section. Second to show the reader that monoids exist in nature. Needless to say, the example will not be used in any way throughout the rest of the book.)

Still other examples. At the end of Chapter III, §4, we shall remark that isomorphism classes of modules over a ring form a monoid under the direct sum. In Chapter XV, §1, we shall consider a monoid consisting of equivalence classes of quadratic forms.

§2. GROUPS

A **group** G is a monoid, such that for every element $x \in G$ there exists an element $y \in G$ such that $xy = yx = e$. Such an element y is called an **inverse** for x. Such an inverse is unique, because if y' is also an inverse for x, then

$$y' = y'e = y'(xy) = (y'x)y = ey = y.$$

We denote this inverse by x^{-1} (or by $-x$ when the law of composition is written additively).

For any positive integer n, we let $x^{-n} = (x^{-1})^n$. Then the usual rules for exponentiation hold for all integers, not only for integers $\geqq 0$ (as we pointed out for monoids in §1). The trivial proofs are left to the reader.

In the definitions of unit elements and inverses, we could also define left units and left inverses (in the obvious way). One can easily prove that these are also units and inverses respectively under suitable conditions. Namely:

Let G be a set with an associative law of composition, let e be a left unit for that law, and assume that every element has a left inverse. Then e is a unit, and each left inverse is also an inverse. In particular, G is a group.

To prove this, let $a \in G$ and let $b \in G$ be such that $ba = e$. Then

$$bab = eb = b.$$

Multiplying on the left by a left inverse for b yields

$$ab = e,$$

or in other words, b is also a right inverse for a. One sees also that a is a left inverse for b. Furthermore,

$$ae = aba = ea = a,$$

whence e is a right unit.

Example. Let G be a group and S a nonempty set. The set of maps $M(S, G)$ is itself a group; namely for two maps f, g of S into G we define fg to be the map such that

$$(fg)(x) = f(x)g(x),$$

and we define f^{-1} to be the map such that $f^{-1}(x) = f(x)^{-1}$. It is then trivial to verify that $M(S, G)$ is a group. If G is commutative, so is $M(S, G)$, and when the law of composition in G is written additively, so is the law of composition in $M(S, G)$, so that we would write $f + g$ instead of fg, and $-f$ instead of f^{-1}.

Example. Let S be a non-empty set. Let G be the set of bijective mappings of S onto itself. Then G is a group, the law of composition being ordinary composition of mappings. The unit element of G is the identity map of S, and the other group properties are trivially verified. The elements of G are called **permutations** of S. We also denote G by Perm(S). For more information on Perm(S) when S is finite, see §5 below.

Example. Let us assume here the basic notions of linear algebra. Let k be a field and V a vector space over k. Let $GL(V)$ denote the set of invertible k-linear maps of V onto itself. Then $GL(V)$ is a group under composition of mappings. Similarly, let k be a field and let $GL(n, k)$ be the set of invertible $n \times n$ matrices with components in k. Then $GL(n, k)$ is a group under the multiplication of matrices. For $n \geqq 2$, this group is not commutative.

Example. The group of automorphisms. We recommend that the reader now refer immediately to §11, where the notion of a category is defined, and where several examples are given. For any object A in a category, its automorphisms form a group denoted by Aut(A). Permutations of a set and the linear automorphisms of a vector space are merely examples of this more general structure.

Example. The set of rational numbers forms a group under addition. The set of non-zero rational numbers forms a group under multiplication. Similar statements hold for the real and complex numbers.

Example. Cyclic groups. The integers **Z** form an additive group. A group is defined to be **cyclic** if there exists an element $a \in G$ such that every element of G (written multiplicatively) is of the form a^n for some integer n. If G is written additively, then every element of a cyclic group is of the form na. One calls a a **cyclic generator**. Thus **Z** is an additive cyclic group with generator 1, and also with generator -1. There are no other generators. Given a positive integer n, the n-th roots of unity in the complex numbers form a cyclic group of order n. In terms of the usual notation, $e^{2\pi i/n}$ is a generator for this group. So is $e^{2\pi ir/n}$

with $r \in \mathbf{Z}$ and r prime to n. A generator for this group is called a **primitive** n-th root of unity.

Example. The direct product. Let G_1, G_2 be groups. Let $G_1 \times G_2$ be the direct product as sets, so $G_1 \times G_2$ is the set of all pairs (x_1, x_2) with $x_i \in G_i$. We define the law of composition componentwise by

$$(x_1, x_2)(y_1, y_2) = (x_1 y_1, x_2 y_2).$$

Then $G_1 \times G_2$ is a group, whose unit element is (e_1, e_2) (where e_i is the unit element of G_i). Similarly, for n groups we define $G_1 \times \cdots \times G_n$ to be the set of n-tuples with $x_i \in G_i$ $(i = 1, \ldots, n)$, and componentwise multiplication. Even more generally, let I be a set, and for each $i \in I$, let G_i be a group. Let $G = \prod G_i$ be the set-theoretic product of the sets G_i. Then G is the set of all families $(x_i)_{i \in I}$ with $x_i \in G_i$. We can define a group structure on G by componentwise multiplication, namely, if $(x_i)_{i \in I}$ and $(y_i)_{i \in I}$ are two elements of G, we define their product to be $(x_i y_i)_{i \in I}$. We define the inverse of $(x_i)_{i \in I}$ to be $(x_i^{-1})_{i \in I}$. It is then obvious that G is a group called the **direct product** of the family.

Let G be a group. A **subgroup** H of G is a subset of G containing the unit element, and such that H is closed under the law of composition and inverse (i.e. it is a submonoid, such that if $x \in H$ then $x^{-1} \in H$). A subgroup is called **trivial** if it consists of the unit element alone. The intersection of an arbitrary non-empty family of subgroups is a subgroup (trivial verification).

Let G be a group and S a subset of G. We shall say that S **generates** G, or that S is a set of **generators** for G, if every element of G can be expressed as a product of elements of S or inverses of elements of S, i.e. as a product $x_1 \cdots x_n$ where each x_i or x_i^{-1} is in S. It is clear that the set of all such products is a subgroup of G (the empty product is the unit element), and is the smallest subgroup of G containing S. Thus S generates G if and only if the smallest subgroup of G containing S is G itself. If G is generated by S, then we write $G = \langle S \rangle$. By definition, a cyclic group is a group which has one generator. Given elements $x_1, \ldots, x_n \in G$, these elements generate a subgroup $\langle x_1, \ldots, x_n \rangle$, namely the set of all elements of G of the form

$$x_{i_1}^{k_1} \cdots x_{i_r}^{k_r} \quad \text{with} \quad k_1, \ldots, k_r \in \mathbf{Z}.$$

A single element $x \in G$ generates a cyclic subgroup.

Example. There are two non-abelian groups of order 8. One is the **group of symmetries of the square**, generated by two elements σ, τ such that

$$\sigma^4 = \tau^2 = e \quad \text{and} \quad \tau\sigma\tau^{-1} = \sigma^3.$$

The other is the **quaternion group**, generated by two elements, i, j such that if we put $k = ij$ and $m = i^2$, then

$$i^4 = j^4 = k^4 = e, \quad i^2 = j^2 = k^2 = m, \quad ij = mji.$$

After you know enough facts about groups, you can easily do Exercise 35.

Let G, G' be monoids. A **monoid-homomorphism** (or simply **homomorphism**) of G into G' is a mapping $f: G \to G'$ such that $f(xy) = f(x)f(y)$ for all $x, y \in G$, and mapping the unit element of G into that of G'. If G, G' are groups, a **group-homomorphism** of G into G' is simply a monoid-homomorphism.

We sometimes say: "Let $f: G \to G'$ be a group-homomorphism" to mean: "Let G, G' be groups, and let f be a homomorphism from G into G'."

Let $f: G \to G'$ be a group-homomorphism. Then

$$f(x^{-1}) = f(x)^{-1}$$

because if e, e' are the unit elements of G, G' respectively, then

$$e' = f(e) = f(xx^{-1}) = f(x)f(x^{-1}).$$

Furthermore, if G, G' are groups and $f: G \to G'$ is a map such that

$$f(xy) = f(x)f(y)$$

for all x, y in G, then $f(e) = e'$ because $f(ee) = f(e)$ and also $= f(e)f(e)$. Multiplying by the inverse of $f(e)$ shows that $f(e) = e'$.

Let G, G' be monoids. A homomorphism $f: G \to G'$ is called an **isomorphism** if there exists a homomorphism $g: G' \to G$ such that $f \circ g$ and $g \circ f$ are the identity mappings (in G' and G respectively). It is trivially verified that f is an isomorphism if and only if f is bijective. The existence of an isomorphism between two groups G and G' is sometimes denoted by $G \approx G'$. If $G = G'$, we say that isomorphism is an **automorphism**. A homomorphism of G into itself is also called an **endomorphism**.

Example. Let G be a monoid and x an element of G. Let $\mathbf{N}$ denote the (additive) monoid of integers $\geqq 0$. Then the map $f: \mathbf{N} \to G$ such that $f(n) = x^n$ is a homomorphism. If G is a group, we can extend f to a homomorphism of $\mathbf{Z}$ into G (x^n is defined for all $n \in \mathbf{Z}$, as pointed out previously). The trivial proofs are left to the reader.

Let n be a fixed integer and let G be a *commutative* group. Then one verifies easily that the map

$$x \mapsto x^n$$

from G into itself is a homomorphism. So is the map $x \mapsto x^{-1}$. The map $x \mapsto x^n$ is called the n-th **power map**.

Example. Let $I = \{i\}$ be an indexing set, and let $\{G_i\}$ be a family of groups. Let $G = \prod G_i$ be their direct product. Let

$$p_i\colon G \to G_i$$

be the projection on the i-th factor. Then p_i is a homomorphism.

Let G be a group, S a set of generators for G, and G' another group. Let $f: S \to G'$ be a map. If there exists a homomorphism $\bar{f}$ of G into G' whose restriction to S is f, then there is only one.

In other words, f has at most one extension to a homomorphism of G into G'. This is obvious, but will be used many times in the sequel.

Let $f: G \to G'$ and $g: G' \to G''$ be two group-homomorphisms. Then the composite map $g \circ f$ is a group-homomorphism. If f, g are isomorphisms then so is $g \circ f$. Furthermore $f^{-1}: G' \to G$ is also an isomorphism. In particular, the set of all automorphisms of G is itself a group, denoted by $\mathrm{Aut}(G)$.

Let $f: G \to G'$ be a group-homomorphism. Let e, e' be the respective unit elements of G, G'. We define the **kernel** of f to be the subset of G consisting of all x such that $f(x) = e'$. From the definitions, it follows at once that the kernel H of f is a subgroup of G. (Let us prove for instance that H is closed under the inverse mapping. Let $x \in H$. Then

$$f(x^{-1})f(x) = f(e) = e'.$$

Since $f(x) = e'$, we have $f(x^{-1}) = e'$, whence $x^{-1} \in H$. We leave the other verifications to the reader.)

Let $f: G \to G'$ be a group-homomorphism again. Let H' be the **image** of f. Then H' is a subgroup of G', because it contains e', and if $f(x), f(y) \in H'$, then $f(xy) = f(x)f(y)$ lies also in H'. Furthermore, $f(x^{-1}) = f(x)^{-1}$ is in H', and hence H' is a subgroup of G'.

The kernel and image of f are sometimes denoted by $\mathrm{Ker}\, f$ and $\mathrm{Im}\, f$.

A homomorphism $f: G \to G'$ which establishes an isomorphism between G and its image in G' will also be called an **embedding**.

A homomorphism whose kernel is trivial is injective.

To prove this, suppose that the kernel of f is trivial, and let $f(x) = f(y)$ for some $x, y \in G$. Multiplying by $f(y^{-1})$ we obtain

$$f(xy^{-1}) = f(x)f(y^{-1}) = e'.$$

Hence xy^{-1} lies in the kernel, hence $xy^{-1} = e$, and $x = y$. If in particular f is also surjective, then f is an isomorphism. Thus a surjective homomorphism whose kernel is trivial must be an isomorphism. We note that an injective homomorphism is an embedding.

An injective homomorphism is often denoted by a special arrow, such as

$$f: G \hookrightarrow G'.$$

There is a useful criterion for a group to be a direct product of subgroups:

Proposition 2.1. *Let G be a group and let H, K be two subgroups such that $H \cap K = e$, $HK = G$, and such that $xy = yx$ for all $x \in H$ and $y \in K$. Then the map*

$$H \times K \to G$$

such that $(x, y) \mapsto xy$ is an isomorphism.

Proof. It is obviously a homomorphism, which is surjective since $HK = G$.

If (x, y) is in its kernel, then $x = y^{-1}$, whence x lies in both H and K, and $x = e$, so that $y = e$ also, and our map is an isomorphism.

We observe that Proposition 2.1 generalizes by induction to a finite number of subgroups $H_1, \dots, H_n$ whose elements commute with each other, such that

$$H_1 \cdots H_n = G,$$

and such that

$$H_{i+1} \cap (H_1 \cdots H_i) = e.$$

In that case, G is isomorphic to the direct product

$$H_1 \times \cdots \times H_n.$$

Let G be a group and H a subgroup. A **left coset** of H in G is a subset of G of type aH, for some element a of G. An element of aH is called a **coset representative** of aH. The map $x \mapsto ax$ induces a bijection of H onto aH. Hence any two left cosets have the same cardinality.

Observe that if a, b are elements of G and aH, bH are cosets having one element in common, then they are equal. Indeed, let $ax = by$ with $x, y \in H$. Then $a = byx^{-1}$. But $yx^{-1} \in H$. Hence $aH = b(yx^{-1})H = bH$, because for any $z \in H$ we have $zH = H$.

We conclude that G is the disjoint union of the left cosets of H. A similar remark applies to **right cosets** (i.e. subsets of G of type Ha). The number of left cosets of H in G is denoted by $(G:H)$, and is called the (left) **index** of H in G. The index of the trivial subgroup is called the **order** of G and is written $(G:1)$. From the above conclusion, we get:

Proposition 2.2. *Let G be a group and H a subgroup. Then*

$$(G:H)(H:1) = (G:1),$$

in the sense that if two of these indices are finite, so is the third and equality holds as stated. If $(G:1)$ is finite, the order of H divides the order of G.

More generally, let H, K be subgroups of G and let $H \supset K$. Let $\{x_i\}$ be a set of (left) coset representatives of K in H and let $\{y_j\}$ be a set of coset representatives of H in G. Then we contend that $\{y_j x_i\}$ is a set of coset representatives of K in G.

Proof. Note that

$$H = \bigcup_i x_i K \qquad \text{(disjoint)},$$

$$G = \bigcup_j y_j H \qquad \text{(disjoint)}.$$

Hence

$$G = \bigcup_{i,j} y_j x_i K.$$

We must show that this union is disjoint, i.e. that the $y_j x_i$ represent distinct cosets. Suppose

$$y_j x_i K = y_{j'} x_{i'} K$$

for a pair of indices (j, i) and (j', i'). Multiplying by H on the right, and noting that x_i, $x_{i'}$ are in H, we get

$$y_j H = y_{j'} H,$$

whence $y_j = y_{j'}$. From this it follows that $x_i K = x_{i'} K$ and therefore that $x_i = x_{i'}$, as was to be shown.

The formula of Proposition 2.2 may therefore be generalized by writing

$$(G : K) = (G : H)(H : K),$$

with the understanding that if two of the three indices appearing in this formula are finite, then so is the third and the formula holds.

The above results are concerned systematically with left cosets. For the right cosets, see Exercise 10.

Example. A group of prime order is cyclic. Indeed, let G have order p and let $a \in G$, $a \neq e$. Let H be the subgroup generated by a. Then $\#(H)$ divides p and is $\neq 1$, so $\#(H) = p$ and so $H = G$, which is therefore cyclic.

Example. Let $J_n = \{1, \dots, n\}$. Let S_n be the group of permutations of J_n. We define a **transposition** to be a permutation τ such that there exist two elements $r \neq s$ in J_n for which $\tau(r) = s$, $\tau(s) = r$, and $\tau(k) = k$ for all $k \neq r, s$. Note that the transpositions generate S_n. Indeed, say σ is a permutation, $\sigma(n) = k \neq n$. Let τ be the transposition interchanging k, n. Then $\tau\sigma$ leaves n fixed, and by induction, we can write $\tau\sigma$ as a product of transpositions in $\mathrm{Perm}(J_{n-1})$, thus proving that transpositions generate S_n.

Next we note that $\#(S_n) = n!$. Indeed, let H be the subgroup of S_n consisting of those elements which leave n fixed. Then H may be identified with S_{n-1}. If σ_i $(i = 1, \dots, n)$ is an element of S_n such that $\sigma_i(n) = i$, then it is immediately verified that $\sigma_1, \dots, \sigma_n$ are coset representatives of H. Hence by induction

$$(S_n : 1) = n(H : 1) = n!.$$

Observe that for σ_i we could have taken the transposition τ_i, which interchanges i and n (except for $i = n$, where we could take σ_n to be the identity).

§3. NORMAL SUBGROUPS

We have already observed that the kernel of a group-homomorphism is a subgroup. We now wish to characterize such subgroups.

Let $f : G \to G'$ be a group-homomorphism, and let H be its kernel. If x is an element of G, then $xH = Hx$, because both are equal to $f^{-1}(f(x))$. We can also rewrite this relation as $xHx^{-1} = H$.

Conversely, let G be a group, and let H be a subgroup. Assume that for all elements x of G we have $xH \subset Hx$ (or equivalently, $xHx^{-1} \subset H$). If we write x^{-1} instead of x, we get $H \subset xHx^{-1}$, whence $xHx^{-1} = H$. Thus our condition is equivalent to the condition $xHx^{-1} = H$ for all $x \in G$. A subgroup H satisfying this condition will be called **normal**. We shall now see that a normal subgroup is the kernel of a homomorphism.

Let G' be the set of cosets of H. (By assumption, a left coset is equal to a right coset, so we need not distinguish between them.) If xH and yH are cosets, then their product $(xH)(yH)$ is also a coset, because

$$xHyH = xyHH = xyH.$$

By means of this product, we have therefore defined a law of composition on G' which is associative. It is clear that the coset H itself is a unit element for this law of composition, and that $x^{-1}H$ is an inverse for the coset xH. Hence G' is a group.

Let $f: G \to G'$ be the mapping such that $f(x)$ is the coset xH. Then f is clearly a homomorphism, and (the subgroup) H is contained in its kernel. If $f(x) = H$, then $xH = H$. Since H contains the unit element, it follows that $x \in H$. Thus H is equal to the kernel, and we have obtained our desired homomorphism.

The group of cosets of a normal subgroup H is denoted by G/H (which we read G modulo H, or G mod H). The map f of G onto G/H constructed above is called the **canonical map**, and G/H is called the **factor group** of G by H.

Remarks

1. Let $\{H_i\}_{i \in I}$ be a family of normal subgroups of G. Then the subgroup

$$H = \bigcap_{i \in I} H_i$$

is a normal subgroup. Indeed, if $y \in H$, and $x \in G$, then xyx^{-1} lies in each H_i, whence in H.

2. Let S be a subset of G and let $N = N_S$ be the set of all elements $x \in G$ such that $xSx^{-1} = S$. Then N is obviously a subgroup of G, called the **normalizer** of S. If S consists of one element a, then N is also called the **centralizer** of a. More generally, let Z_S be the set of all elements $x \in G$ such that $xyx^{-1} = y$ for all $y \in S$. Then Z_S is called the **centralizer** of S. The centralizer of G itself is called the **center** of G. It is the subgroup of G consisting of all elements of G commuting with all other elements, and is obviously a normal subgroup of G.

Examples. We shall give more examples of normal subgroups later when we have more theorems to prove the normality. Here we give only two examples.

First, from linear algebra, note that the determinant is a homomorphism from the multiplicative group of square matrices into the multiplicative group of a field. The kernel is called the **special linear group**, and is normal.

Second, let G be the set of all maps $T_{a,b}\colon \mathbf{R} \to \mathbf{R}$ such that $T_{a,b}(x) = ax + b$, with $a \neq 0$ and b arbitrary. Then G is a group under composition of mappings. Let A be the multiplicative group of maps of the form $T_{a,0}$ (isomorphic to $\mathbf{R}^*$, the non-zero elements of $\mathbf{R}$), and let N be the group of translations $T_{1,b}$ with $b \in \mathbf{R}$. Then the reader will verify at once that $T_{a,b} \mapsto a$ is a homomorphism of G onto the multiplicative group, whose kernel is the group of translations, which is therefore normal. Furthermore, we have $G = AN = NA$, and $N \cap A = \{\mathrm{id}\}$. In the terminology of Exercise 12, G is the **semidirect product** of A and N.

Let H be a subgroup of G. Then H is obviously a normal subgroup of its normalizer N_H. We leave the following statements as exercises:

If K is any subgroup of G containing H and such that H is normal in K, then $K \subset N_H$.

If K is a subgroup of N_H, then KH is a group and H is normal in KH.
The normalizer of H is the largest subgroup of G in which H is normal.

Let G be a group and H a normal subgroup. Let $x, y \in G$. We shall write

$$x \equiv y \pmod{H}$$

if x and y lie in the same coset of H, or equivalently if xy^{-1} (or $y^{-1}x$) lie in H. We read this relation "x and y are congruent modulo H."

When G is an additive group, then

$$x \equiv 0 \pmod{H}$$

means that x lies in H, and

$$x \equiv y \pmod{H}$$

means that $x - y$ (or $y - x$) lies in H. This notation of congruence is used mostly for additive groups.

Let

$$G' \xrightarrow{f} G \xrightarrow{g} G''$$

be a sequence of homomorphisms. We shall say that this sequence is **exact** if $\operatorname{Im} f = \operatorname{Ker} g$. For example, if H is a normal subgroup of G then the sequence

$$H \xrightarrow{j} G \xrightarrow{\varphi} G/H$$

is exact (where j = inclusion and φ = canonical map). A sequence of homomorphisms having more than one term, like

$$G_1 \xrightarrow{f_1} G_2 \xrightarrow{f_2} G_3 \to \cdots \xrightarrow{f_{n-1}} G_n,$$

is called **exact** if it is exact at each joint, i.e. if for each $i = 1, \dots, n-2$,

$$\operatorname{Im} f_i = \operatorname{Ker} f_{i+1}.$$

For example letting 0 be the trivial group, to say that

$$0 \to G' \xrightarrow{f} G \xrightarrow{g} G'' \to 0$$

is exact means that f is injective, that $\operatorname{Im} f = \operatorname{Ker} g$, and that g is surjective. If $H = \operatorname{Ker} g$ then this sequence is essentially the same as the exact sequence

$$0 \to H \to G \to G/H \to 0.$$

More precisely, there exists a commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & G' \arrow[r, "f"] \arrow[d] & G \arrow[r, "g"] \arrow[d] & G'' \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & H \arrow[r] & G \arrow[r] & G/H \arrow[r] & 0
\end{tikzcd}$$

in which the vertical maps are isomorphisms, and the rows are exact.

Next we describe some homomorphisms, all of which are called **canonical**.

(i) Let G, G' be groups and $f: G \to G'$ a homomorphism whose kernel is H. Let $\varphi: G \to G/H$ be the canonical map. Then there exists a unique homomorphism $f_*: G/H \to G'$ such that $f = f_* \circ \varphi$, and f_* is injective.

To define f_*, let xH be a coset of H. Since $f(xy) = f(x)$ for all $y \in H$, we define $f_*(xH)$ to be $f(x)$. This value is independent of the choice of coset representative x, and it is then trivially verified that f_* is a homomorphism, is injective, and is the unique homomorphism satisfying our requirements. We shall say that f_* is **induced** by f.

Our homomorphism f_* induces an isomorphism

$$\lambda: G/H \to \operatorname{Im} f$$

of G/H onto the image of f, and thus f can be factored into the following succession of homomorphisms:

$$G \xrightarrow{\varphi} G/H \xrightarrow{\lambda} \operatorname{Im} f \xrightarrow{j} G'.$$

Here, j is the inclusion of $\operatorname{Im} f$ in G'.

(ii) Let G be a group and H a subgroup. Let N be the intersection of all normal subgroups containing H. Then N is normal, and hence is the smallest normal subgroup of G containing H. Let $f: G \to G'$ be a homomorphism whose kernel contains H. Then the kernel of f contains N, and there exists a unique homomorphism $f_*: G/N \to G'$, said to be induced by f, making the following diagram commutative:

$$\begin{tikzcd}
G \arrow[rr, "f"] \arrow[dr, "\varphi"'] & & G' \\
& G/N \arrow[ur, "f_*"'] &
\end{tikzcd}$$

As before, φ is the canonical map.

We can define f_* as in (1) by the rule

$$f_*(xN) = f(x).$$

This is well defined, and is trivially verified to satisfy all our requirements.

(iii) Let G be group and $H \supset K$ two normal subgroups of G. Then K is normal in H, and we can define a map of G/K onto G/H by associating with each coset xK the coset xH. It is immediately verified that this map is a homomorphism, and that its kernel consists of all cosets xK such that $x \in H$. Thus we have a canonical isomorphism

$$\boxed{(G/K)/(H/K) \approx G/H.}$$

One could also describe this isomorphism using (i) and (ii). We leave it to the reader to show that we have a commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & H \arrow[r] \arrow[d, "\text{can}"] & G \arrow[r] \arrow[d, "\text{can}"] & G/H \arrow[r] \arrow[d, "\text{id}"] & 0 \\
0 \arrow[r] & H/K \arrow[r] & G/K \arrow[r] & G/H \arrow[r] & 0
\end{tikzcd}$$

where the rows are exact.

(iv) Let G be a group and let H, K be two subgroups. Assume that H is contained in the normalizer of K. Then $H \cap K$ is obviously a normal subgroup of H, and equally obviously $HK = KH$ is a subgroup of G. There is a surjective homomorphism

$$H \to HK/K$$

associating with each $x \in H$ the coset xK of K in the group HK. The reader will verify at once that the kernel of this homomorphism is exactly $H \cap K$. Thus we have a canonical isomorphism

$$\boxed{H/(H \cap K) \approx HK/K.}$$

(v) Let $f\colon G \to G'$ be a group homomorphism, let H' be a normal subgroup of G', and let $H = f^{-1}(H')$.

$$\begin{tikzcd}
G \arrow[r] & G' \\
f^{-1}(H') \arrow[r] \arrow[u] & H' \arrow[u]
\end{tikzcd}$$

Then $f^{-1}(H')$ is normal in G. [*Proof*: If $x \in G$, then $f(xHx^{-1}) = f(x)f(H)f(x)^{-1}$ is contained in H', so $xHx^{-1} \subset H$.] We then obtain a homomorphism

$$G \to G' \to G'/H'$$

composing f with the canonical map of G' onto G'/H', and the kernel of this composite is H. Hence we get an injective homomorphism

$$\bar{f}\colon G/H \to G'/H'$$

again called canonical, giving rise to the commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & H \arrow[r] \arrow[d] & G \arrow[r] \arrow[d, "f"] & G/H \arrow[r] \arrow[d, "\bar{f}"] & 0 \\
0 \arrow[r] & H' \arrow[r] & G' \arrow[r] & G'/H' \arrow[r] & 0.
\end{tikzcd}$$

If f is surjective, then $\bar{f}$ is an isomorphism.

We shall now describe some applications of our homomorphism statements.

Let G be a group. A sequence of subgroups

$$G = G_0 \supset G_1 \supset G_2 \supset \cdots \supset G_m$$

is called a **tower** of subgroups. The tower is said to be **normal** if each G_{i+1} is normal in G_i $(i = 0, \ldots, m - 1)$. It is said to be **abelian** (resp. **cyclic**) if it is normal and if each factor group G_i/G_{i+1} is abelian (resp. cyclic).

Let $f \colon G \to G'$ be a homomorphism and let

$$G' = G'_0 \supset G'_1 \supset \cdots \supset G'_m$$

be a normal tower in G'. Let $G_i = f^{-1}(G'_i)$. Then the G_i $(i = 0, \ldots, m)$ form a normal tower. If the G'_i form an abelian tower (resp. cyclic tower) then the G_i form an abelian tower (resp. cyclic tower), because we have an injective homomorphism

$$G_i/G_{i+1} \to G'_i/G'_{i+1}$$

for each i, and because a subgroup of an abelian group (resp. a cyclic group) is abelian (resp. cyclic).

A **refinement** of a tower

$$G = G_0 \supset G_1 \supset \cdots \supset G_m$$

is a tower which can be obtained by inserting a finite number of subgroups in the given tower. A group is said to be **solvable** if it has an abelian tower, whose last element is the trivial subgroup (i.e. $G_m = \{e\}$ in the above notation).

Proposition 3.1. *Let G be a finite group. An abelian tower of G admits a cyclic refinement. Let G be a finite solvable group. Then G admits a cyclic tower whose last element is $\{e\}$.*

Proof. The second assertion is an immediate consequence of the first, and it clearly suffices to prove that if G is finite, abelian, then G admits a cyclic tower ending with $\{e\}$. We use induction on the order of G. Let x be an element of G. We may assume that $x \neq e$. Let X be the cyclic group generated by x. Let $G' = G/X$. By induction, we can find a cyclic tower in G', and its inverse image is a cyclic tower in G whose last element is X. If we refine this tower by inserting $\{e\}$ at the end, we obtain the desired cyclic tower.

Example. In Theorem 6.5 it will be proved that a group whose order is a prime power is solvable.

Example. One of the major results of group theory is the Feit-Thompson theorem that all finite groups of odd order are solvable. Cf. [Go 68].

Example. Solvable groups will occur in field theory as the Galois groups of solvable extensions. See Chapter VI, Theorem 7.2.

Example. We assume the reader knows the basic notions of linear algebra. Let k be a field. Let $G = GL(n, k)$ be the group of invertible $n \times n$ matrices in k. Let $T = T(n, k)$ be the upper triangular group; that is, the subgroup of matrices which are 0 below the diagonal. Let D be the diagonal group of diagonal matrices with non-zero components on the diagonal. Let N be the additive group of matrices which are 0 on and below the diagonal, and let $U = I + N$, where I is the unit $n \times n$ matrix. Then U is a subgroup of G. (Note that N consists of nilpotent matrices, i.e. matrices A such that $A^m = 0$ for some positive integer m. Then $(I - A)^{-1} = I + A + A^2 + \ldots + A^{m-1}$ is computed using the geometric series.) Given a matrix $A \in T$, let $\operatorname{diag}(A)$ be the diagonal matrix which has the same diagonal components as A. Then the reader will verify that we get a surjective homomorphism

$$T \to D \quad \text{given by} \quad A \mapsto \operatorname{diag}(A).$$

The kernel of this homomorphism is precisely U. More generally, observe that for $r \geqq 2$, the set N^{r-1} consists of all matrices of the form

$$M = \begin{pmatrix} 0 & 0 & \cdots & 0 & a_{1r} & & \cdots\cdots & a_{1n} \\ 0 & 0 & \cdots & 0 & 0 & a_{2,r+1} & \cdots & a_{2n} \\ \vdots & \vdots & & & & & \ddots & \vdots \\ 0 & 0 & \cdots\cdots & & & & & a_{n-r+1,n} \\ 0 & 0 & \cdots\cdots & & & & & 0 \\ & & & & \cdots & & & \\ 0 & 0 & \cdots\cdots & & & & & 0 \end{pmatrix}$$

Let $U_r = I + N^r$. Then $U_1 = U$ and $U_r \supset U_{r+1}$. Furthermore, U_{r+1} is normal in U_r, and the factor group is isomorphic to the additive group (!) k^{n-r}, under the the mapping which sends $I + M$ to the $n - r$-tuple $(a_{1r+1}, \ldots, a_{n-r,n}) \in k^{n-r}$. This $n - r$-tuple could be called the r-th upper diagonal. Thus we obtain an abelian tower

$$T \supset U = U_1 \supset U_2 \supset \ldots \supset U_n = \{I\}.$$

Theorem 3.2. *Let G be a group and H a normal subgroup. Then G is solvable if and only if H and G/H are solvable.*

Proof. We prove that G solvable implies that H is solvable. Let $G = G_0 \supset G_1 \supset \ldots \supset G_r = \{e\}$ be a tower of groups with G_{i+1} normal in G_i and such that G_i/G_{i+1} is abelian. Let $H_i = H \cap G_i$. Then H_{i+1} is normal in H_i, and we have an embedding $H_i/H_{i+1} \to G_i/G_{i+1}$, whence H_i/H_{i+1} is abelian, whence proving that H is solvable. We leave the proofs of the other statements to the reader.

Let G be a group. A **commutator** in G is a group element of the form $xyx^{-1}y^{-1}$ with $x, y \in G$. Let G^c be the subgroup of G generated by the commutators. We call G^c the **commutator subgroup** of G. As an exercise, prove that G^c is normal in G, and that every homomorphism $f : G \to G'$ into a commutative group G' contains G^c in its kernel, and consequently factors through the factor commutator group G/G^c. Observe that G/G^c itself is commutative. Indeed, if $\bar{x}$ denotes the image of x in G/G^c, then by definition we have $\bar{x}\bar{y}\bar{x}^{-1}\bar{y}^{-1} = \bar{e}$, so $\bar{x}$ and $\bar{y}$ commute. In light of the definition of solvability, it is clear that the commutator group is at the heart of solvability and non-solvability problems.

A group G is said to be **simple** if it is non-trivial, and has no normal subgroups other than $\{e\}$ and G itself.

Examples. An abelian group is simple if and only if it is cyclic of prime order. Indeed, suppose A abelian and non-trivial. Let $a \in A$, $a \neq e$. If a generates an infinite cyclic group, then a^2 generates a proper subgroup and so A is not simple. If a has finite period, and A is simple, then $A = \langle a \rangle$. Let n be the period and suppose n not prime. Write $n = rs$ with $r, s > 1$. Then $a^r \neq e$ and a^r generates a proper subgroup, contradicting the simplicity of A, so a has prime period and A is cyclic of order p.

Examples. Using commutators, we shall give examples of simple groups in Theorem 5.5 (the alternating group), and in Theorem 9.2 of Chapter XIII ($PSL_n(F)$, a group of matrices to be defined in that chapter). Since a non-cyclic simple group is not solvable, we get thereby examples of non-solvable groups.

A major program of finite group theory is the classification of all finite simple groups. Essentially most of them (if not all) have natural representations as subgroups of linear maps of suitable vector spaces over suitable fields, in a suitably natural way. See [Go 82], [Go 86], [Sol 01] for surveys. Gaps in purported proofs have been found. As of 2001, these are still incomplete.

Next we are concerned with towers of subgroups such that the factor groups G_i/G_{i+1} are simple. The next lemma is for use in the proof of the Jordan-Hölder and Schreier theorems.

Lemma 3.3. (Butterfly Lemma.) (Zassenhaus) *Let U, V be subgroups of a group. Let u, v be normal subgroups of U and V, respectively. Then*

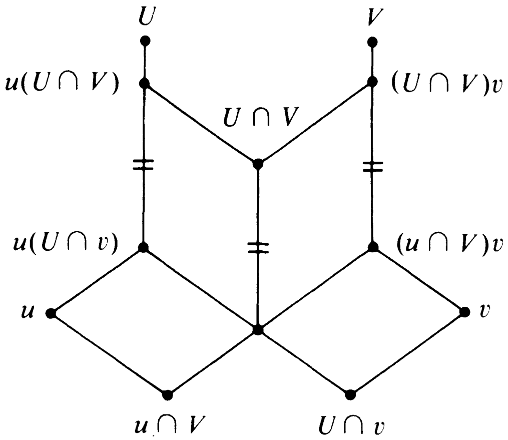
$$u(U \cap v) \text{ is normal in } u(U \cap V),$$

$$(u \cap V)v \text{ is normal in } (U \cap V)v,$$

and the factor groups are isomorphic, i.e.

$$u(U \cap V)/u(U \cap v) \approx (U \cap V)v/(u \cap V)v.$$

Proof. The combination of groups and factor groups becomes clear if one visualizes the following diagram of subgroups (which gives its name to the lemma):



In this diagram, we are given U, u, V, v. All the other points in the diagram correspond to certain groups which can be determined as follows. The intersection of two line segments going downwards represents the intersection of groups. Two lines going upwards meet in a point which represents the product of two subgroups (i.e. the smallest subgroup containing both of them).

We consider the two parallelograms representing the wings of the butterfly, and we shall give isomorphisms of the factor groups as follows:

$$\frac{u(U \cap V)}{u(U \cap v)} \approx \frac{U \cap V}{(u \cap V)(U \cap v)} \approx \frac{(U \cap V)v}{(u \cap V)v}.$$

In fact, the vertical side common to both parallelograms has $U \cap V$ as its top end point, and $(u \cap V)(U \cap v)$ as its bottom end point. We have an isomorphism

$$(U \cap V)/(u \cap V)(U \cap v) \approx u(U \cap V)/u(U \cap v).$$

This is obtained from the isomorphism theorem

$$H/(H \cap N) \approx HN/N$$

by setting $H = U \cap V$ and $N = u(U \cap v)$. This gives us the isomorphism on the left. By symmetry we obtain the corresponding isomorphism on the right, which proves the Butterfly lemma.

Let G be a group, and let

$$G = G_1 \supset G_2 \supset \cdots \supset G_r = \{e\},$$

$$G = H_1 \supset H_2 \supset \cdots \supset H_s = \{e\}$$

be normal towers of subgroups, ending with the trivial group. We shall say that these towers are **equivalent** if $r = s$ and if there exists a permutation of the indices $i = 1, \dots, r - 1$, written $i \mapsto i'$, such that

$$G_i/G_{i+1} \approx H_{i'}/H_{i'+1}.$$

In other words, the sequences of factor groups in our two towers are the same, up to isomorphisms, and a permutation of the indices.

Theorem 3.4. (Schreier) *Let G be a group. Two normal towers of subgroups ending with the trivial group have equivalent refinements.*

Proof. Let the two towers be as above. For each $i = 1, \dots, r - 1$ and $j = 1, \dots, s$ we define

$$G_{ij} = G_{i+1}(H_j \cap G_i).$$

Then $G_{is} = G_{i+1}$, and we have a refinement of the first tower:

$$\begin{aligned} G = G_{11} \supset G_{12} \supset \cdots \supset G_{1,s-1} \supset G_2 \\ = G_{21} \supset G_{22} \supset \cdots \supset G_{r-1,1} \supset \cdots \supset G_{r-1,s-1} \supset \{e\}. \end{aligned}$$

Similarly, we define

$$H_{ji} = H_{j+1}(G_i \cap H_j),$$

for $j = 1, \dots, s - 1$ and $i = 1, \dots, r$. This yields a refinement of the second tower. By the butterfly lemma, for $i = 1, \dots, r - 1$ and $j = 1, \dots, s - 1$ we have isomorphisms

$$G_{ij}/G_{i,j+1} \approx H_{ji}/H_{j,i+1}.$$

We view each one of our refined towers as having $(r - 1)(s - 1) + 1$ elements, namely G_{ij} $(i = 1, \dots, r - 1; j = 1, \dots, s - 1)$ and $\{e\}$ in the first case, H_{ji} and $\{e\}$ in the second case. The preceding isomorphism for each pair of indices (i, j) shows that our refined towers are equivalent, as was to be proved.

A group G is said to be **simple** if it is non-trivial, and has no normal subgroups other than $\{e\}$ and G itself.

Theorem 3.5. (Jordan-Hölder) *Let G be a group, and let*

$$G = G_1 \supset G_2 \supset \cdots \supset G_r = \{e\}$$

be a normal tower such that each group G_i/G_{i+1} is simple, and $G_i \neq G_{i+1}$ for $i = 1, \dots, r - 1$. Then any other normal tower of G having the same properties is equivalent to this one.

Proof. Given any refinement $\{G_{ij}\}$ as before for our tower, we observe that for each i, there exists precisely one index j such that $G_i/G_{i+1} = G_{ij}/G_{i,j+1}$. Thus the sequence of non-trivial factors for the original tower, or the refined tower, is the same. This proves our theorem.

Bibliography

[Go 68] D. GORENSTEIN, *Finite groups*, Harper and Row, 1968

[Go 82] D. GORENSTEIN, *Finite simple groups*, Plenum Press, 1982

[Go 83] D. GORENSTEIN, *The Classification of Finite Simple Groups*, Plenum Press, 1983

[Go 86] D. GORENSTEIN, Classifying the finite simple groups, *Bull. AMS* **14** No. 1 (1986), pp. 1–98

[So 01] R. SOLOMON, A brief history of the classification of the finite simple groups, *Bull. AMS* **38**, 3 (2001) pp. 315–352

§4. CYCLIC GROUPS

The integers $\mathbf{Z}$ form an additive group. We shall determine its subgroups. Let H be a subgroup of $\mathbf{Z}$. If H is not trivial, let a be the smallest positive integer in H. We contend that H consists of all elements na, with $n \in \mathbf{Z}$. To prove this, let $y \in H$. There exist integers n, r with $0 \leqq r < a$ such that

$$y = na + r.$$

Since H is a subgroup and $r = y - na$, we have $r \in H$, whence $r = 0$, and our assertion follows.

Let G be a group. We shall say that G is **cyclic** if there exists an element a of G such that every element x of G can be written in the form a^n for some $n \in \mathbf{Z}$ (in other words, if the map $f: \mathbf{Z} \to G$ such that $f(n) = a^n$ is surjective). Such an element a of G is then called a **generator** of G.

Let G be a group and $a \in G$. The subset of all elements a^n $(n \in \mathbf{Z})$ is obviously a cyclic subgroup of G. If m is an integer such that $a^m = e$ and $m > 0$ then we shall call m an **exponent** of a. We shall say that $m > 0$ is an **exponent** of G if $x^m = e$ for all $x \in G$.

Let G be a group and $a \in G$. Let $f: \mathbf{Z} \to G$ be the homomorphism such that $f(n) = a^n$ and let H be the kernel of f. Two cases arise:

1. The kernel is trivial. Then f is an isomorphism of $\mathbf{Z}$ onto the cyclic subgroup of G generated by a, and this subgroup is infinite cyclic. If a generates G, then G is cyclic. We also say that a has **infinite period**.

2. The kernel is not trivial. Let d be the smallest positive integer in the kernel. Then d is called the **period** of a. If m is an integer such that $a^m = e$ then $m = ds$ for some integer s. We observe that the elements $e, a, \ldots, a^{d-1}$ are distinct. Indeed, if $a^r = a^s$ with $0 \leqq r, s \leqq d - 1$, and say $r \leqq s$, then $a^{s-r} = e$. Since $0 \leqq s - r < d$ we must have $s - r = 0$. The cyclic subgroup generated by a has order d. Hence by Proposition 2.2:

Proposition 4.1. *Let G be a finite group of order $n > 1$. Let a be an element of G, $a \neq e$. Then the period of a divides n. If the order of G is a prime number p, then G is cyclic and the period of any generator is equal to p.*

Furthermore:

Proposition 4.2. *Let G be a cyclic group. Then every subgroup of G is cyclic. If f is a homomorphism of G, then the image of f is cyclic.*

Proof. If G is infinite cyclic, it is isomorphic to $\mathbf{Z}$, and we determined above all subgroups of $\mathbf{Z}$, finding that they are all cyclic. If $f: G \to G'$ is a homomorphism, and a is a generator of G, then $f(a)$ is obviously a generator of $f(G)$, which is therefore cyclic, so the image of f is cyclic. Next let H be a subgroup of G. We want to show H cyclic. Let a be a generator of G. Then we have a surjective homomorphism $f: \mathbf{Z} \to G$ such that $f(n) = a^n$. The inverse image $f^{-1}(H)$ is a subgroup of $\mathbf{Z}$, and therefore equal to $m\mathbf{Z}$ for some positive integer m. Since f is surjective, we also have a surjective homomorphism $m\mathbf{Z} \to H$. Since $m\mathbf{Z}$ is cyclic (generated additively by m), it follows that H is cyclic, thus proving the proposition.

We observe that two cyclic groups of the same order m are isomorphic. Indeed, if G is cyclic of order m with generator a, then we have a surjective homomorphism $f: \mathbf{Z} \to G$ such that $f(n) = a^n$, and if $k\mathbf{Z}$ is the kernel, with k positive, then we have an isomorphism $\mathbf{Z}/k\mathbf{Z} \approx G$, so $k = m$. If $u: G_1 \to \mathbf{Z}/m\mathbf{Z}$ and $v: G_2 \to \mathbf{Z}/m\mathbf{Z}$ are isomorphisms of two cyclic groups with $\mathbf{Z}/m\mathbf{Z}$, then $v^{-1} \circ u: G_1 \to G_2$ is an isomorphism.

Proposition 4.3.

(i) *An infinite cyclic group has exactly two generators (if a is a generator, then a^{-1} is the only other generator).*

(ii) *Let G be a finite cyclic group of order n, and let x be a generator. The set of generators of G consists of those powers x^ν of x such that ν is relatively prime to n.*

(iii) *Let G be a cyclic group, and let a, b be two generators. Then there exists an automorphism of G mapping a onto b. Conversely, any automorphism of G maps a on some generator of G.*

(iv) *Let G be a cyclic group of order n. Let d be a positive integer dividing n. Then there exists a unique subgroup of G of order d.*

(v) *Let G_1, G_2 be cyclic of orders m, n respectively. If m, n are relatively prime then $G_1 \times G_2$ is cyclic.*

(vi) *Let G be a finite abelian group. If G is not cyclic, then there exists a prime p and a subgroup of G isomorphic to $C \times C$, where C is cyclic of order p.*

Proof. We leave the first three statements to the reader, and prove the others.

(iv) Let $d|n$. Let $m = n/d$. Let $f : \mathbf{Z} \to G$ be a surjective homomorphism. Then $f(m\mathbf{Z})$ is a subgroup of G, and from the isomorphism $\mathbf{Z}/m\mathbf{Z} \approx G/f(m\mathbf{Z})$ we conclude that $f(m\mathbf{Z})$ has index m in G, whence $f(m\mathbf{Z})$ has order d. Conversely, let H be a subgroup of order d. Then $f^{-1}(H) = m\mathbf{Z}$ for some positive integer m, so $H = f(m\mathbf{Z})$, $\mathbf{Z}/m\mathbf{Z} \approx G/H$, so $n = md$, $m = n/d$ and H is uniquely determined.

(v) Let $A = \langle a \rangle$ and $B = \langle b \rangle$ be cyclic groups of orders m, n, relatively prime. Consider the homomorphism $\mathbf{Z} \to A \times B$ such that $k \mapsto (a^k, b^k)$. An element in its kernel must be divisible both by m and n, hence by their product since m, n are relatively prime. Conversely, it is clear that $mn\mathbf{Z}$ is contained in the kernel, so the kernel is $mn\mathbf{Z}$. The image of $\mathbf{Z} \to A \times B$ is surjective by the Chinese remainder theorem. This proves (v). (A reader who does not know the Chinese remainder theorem can see a proof in the more general context of Chapter II, Theorem 2.2.)

(vi) This characterization of cyclic groups is an immediate consequence of the structure theorem which will be proved in §8, because if G is not cyclic, then by Theorem 8.1 and (v) we are reduced to the case when G is a p-group, and by Theorem 8.2 there are at least two factors in the direct product (or sum) decomposition, and each contains a cyclic subgroup of order p, whence G contains their direct product (or sum). Statement **(vi)** is, of course, easier to prove than the full structure theorem, and it is a good exercise for the reader to formulate the simpler arguments which yield **(vi)** directly.

Note. For the group of automorphisms of a cyclic group, see the end of Chapter II, §2.

§5. OPERATIONS OF A GROUP ON A SET

Let G be a group and let S be a set. An **operation** or an **action** of G on S is a homomorphism

$$\pi : G \to \operatorname{Perm}(S)$$

of G into the group of permutations of S. We then call S a G**-set**. We denote the permutation associated with an element $x \in G$ by π_x. Thus the homomorphism is denoted by $x \mapsto \pi_x$. Given $s \in S$, the image of s under the permutation π_x is $\pi_x(s)$. From such an operation we obtain a mapping

$$G \times S \to S,$$

which to each pair (x, s) with $x \in G$ and $s \in S$ associates the element $\pi_x(s)$. We often abbreviate the notation and write simply xs instead of $\pi_x(s)$. With the simpler notation, we have the two properties:

For all $x, y \in G$ and $s \in S$, we have $x(ys) = (xy)s$.
If e is the unit element of G, then $es = s$ for all $s \in S$.

Conversely, if we are given a mapping $G \times S \rightarrow S$, denoted by $(x, s) \mapsto xs$, satisfying these two properties, then for each $x \in G$ the map $s \mapsto xs$ is a permutation of S, which we then denote by $\pi_x(s)$. Then $x \mapsto \pi_x$ is a homomorphism of G into Perm(S). So an operation of G on S could also be defined as a mapping $G \times S \rightarrow S$ satisfying the above two properties. The most important examples of representations of G as a group of permutations are the following.

1. Conjugation. For each $x \in G$, let $\mathbf{c}_x\colon G \rightarrow G$ be the map such that $\mathbf{c}_x(y) = xyx^{-1}$. Then it is immediately verified that the association $x \mapsto \mathbf{c}_x$ is a homomorphism $G \rightarrow \mathrm{Aut}(G)$, and so this map gives an operation of G on itself, called **conjugation**. The kernel of the homomorphism $x \mapsto \mathbf{c}_x$ is a normal subgroup of G, which consists of all $x \in G$ such that $xyx^{-1} = y$ for all $y \in G$, i.e. all $x \in G$ which commute with every element of G. This kernel is called the **center** of G. Automorphisms of G of the form $\mathbf{c}_x$ are called **inner**.

To avoid confusion about the operation on the left, we don't write xy for $\mathbf{c}_x(y)$. Sometimes, one writes

$$\mathbf{c}_{x^{-1}}(y) = x^{-1}yx = y^x,$$

i.e. one uses an exponential notation, so that we have the rules

$$y^{(xz)} = (y^x)^z \quad \text{and} \quad y^e = y$$

for all $x, y, z \in G$. Similarly, ${}^xy = xyx^{-1}$ and ${}^z({}^xy) = {}^{zx}y$.

We note that G also operates by conjugation on the set of subsets of G. Indeed, let S be the set of subsets of G, and let $A \in S$ be a subset of G. Then xAx^{-1} is also a subset of G which may be denoted by $\mathbf{c}_x(A)$, and one verifies trivially that the map

$$(x, A) \mapsto xAx^{-1}$$

of $G \times S \rightarrow S$ is an operation of G on S. We note in addition that if A is a subgroup of G then xAx^{-1} is also a subgroup, so that G operates on the set of subgroups by conjugation.

If A, B are two subsets of G, we say that they are **conjugate** if there exists $x \in G$ such that $B = xAx^{-1}$.

2. Translation. For each $x \in G$ we define the translation $T_x\colon G \rightarrow G$ by $T_x(y) = xy$. Then the map

$$(x, y) \mapsto xy = T_x(y)$$

defines an operation of G on itself. *Warning*: T_x is not a group-homomorphism! Only a permutation of G.

Similarly, G operates by translation on the set of subsets, for if A is a subset of G, then $xA = T_x(A)$ is also a subset. If H is a subgroup of G, then $T_x(H) = xH$ is in general not a subgroup but a coset of H, and hence we see that G operates by translation on the set of cosets of H. We denote the set of left cosets of H by G/H. Thus even though H need not be normal, G/H is a G-set. It has become customary to denote the set of *right* cosets by $H\backslash G$.

The above two representations of G as a group of permutations will be used frequently in the sequel. In particular, the representation by conjugation will be used throughout the next section, in the proof of the Sylow theorems.

3. Example from linear algebra. We assume the reader knows basic notions of linear algebra. Let k be a field and let V be a vector space over k. Let $G = GL(V)$ be the group of linear automorphisms of V. For $A \in G$ and $v \in V$, the map $(A, v) \mapsto Av$ defines an operation of G on V. Of course, G is a subgroup of the group of permutations Perm(V). Similarly, let $V = k^n$ be the vector space of (vertical) n-tuples of elements of k, and let G be the group of invertible $n \times n$ matrices with components in k. Then G operates on k^n by $(A, X) \mapsto AX$ for $A \in G$ and $X \in k^n$.

Let S, S' be two G-sets, and $f : S \to S'$ a map. We say that f is a **morphism of G-sets**, or a **G-map**, if

$$f(xs) = xf(s)$$

for all $x \in G$ and $s \in S$. (We shall soon define categories, and see that G-sets form a category.)

We now return to the general situation, and consider a group operating on a set S. Let $s \in S$. The set of elements $x \in G$ such that $xs = s$ is obviously a subgroup of G, called the **isotropy** group of s in G, and denoted by G_s.

When G operates on itself by conjugation, then the isotropy group of an element is none other than the normalizer of this element. Similarly, when G operates on the set of subgroups by conjugation, the isotropy group of a subgroup is again its normalizer.

Let G operate on a set S. Let s, s' be elements of S, and y an element of G such that $ys = s'$. Then

$$G_{s'} = yG_s y^{-1}$$

Indeed, one sees at once that yG_sy^{-1} leaves s' fixed. Conversely, if $x's' = s'$ then $x'ys = ys$, so $y^{-1}x'y \in G_s$ and $x' \in yG_sy^{-1}$. Thus the isotropy groups of s and s' are conjugate.

Let K be the kernel of the representation $G \to \text{Perm}(S)$. Then directly from the definitions, we obtain that

$$K = \bigcap_{s \in S} G_s = \text{intersection of all isotropy groups.}$$

An action or operation of G is said to be **faithful** if $K = \{e\}$; that is, the kernel of $G \to \mathrm{Perm}(S)$ is trivial. A **fixed point** of G is an element $s \in S$ such that $xs = s$ for all $x \in G$ or in other words, $G = G_s$.

Let G operate on a set S. Let $s \in S$. The subset of S consisting of all elements xs (with $x \in G$) is denoted by Gs, and is called the **orbit** of s under G. If x and y are in the same coset of the subgroup $H = G_s$, then $xs = ys$, and conversely (obvious). In this manner, we get a mapping

$$f: G/H \to S$$

given by $f(xH) = xs$, and it is clear that this map is a morphism of G-sets. In fact, one sees at once that it induces a bijection of G/H onto the orbit Gs. Consequently:

Proposition 5.1. *If G is a group operating on a set S, and $s \in S$, then the order of the orbit Gs is equal to the index $(G : G_s)$.*

In particular, when G operates by conjugation on the set of subgroups, and H is a subgroup, then:

Proposition 5.2. *The number of conjugate subgroups to H is equal to the index of the normalizer of H.*

Example. Let G be a group and H a subgroup of index 2. Then H is normal in G.

Proof. Note that H is contained in its normalizer N_H, so the index of N_H in G is 1 or 2. If it is 1, then we are done. Suppose it is 2. Let G operate by conjugation on the set of subgroups. The orbit of H has 2 elements, and G operates on this orbit. In this way we get a homomorphism of G into the group of permutations of 2 elements. Since there is one conjugate of H unequal to H, then the kernel of our homomorphism is normal, of index 2, hence equal to H, which is normal, a contradiction which concludes the proof.

For a generalization and other examples, see Lemma 6.7.

In general, an operation of G on S is said to be **transitive** if there is only one orbit.

Examples. The symmetric group S_n operates transitively on $\{1, 2, \ldots, n\}$. (See p. 30.) In Proposition 2.1 of Chapter VII, we shall see a non-trivial example of transitive action of a Galois group operating on the primes lying above a given prime in the ground ring. In topology, suppose we have a universal covering space $p : X' \to X$, where X is connected. Given $x \in X$, the fundamental group $\pi_1(X)$ operates transitively on the inverse image $p^{-1}(x)$.

Example. Let $\mathfrak{H}$ be the upper half-plane; that is, the set of complex numbers $z = x + iy$ such that $y > 0$. Let $G = SL_2(\mathbf{R})$ (2×2 matrices with determinant 1). For

$$\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G, \text{ we let } \alpha z = \frac{az + b}{cz + d}.$$

Readers will verify by brute force that this defines an operation of G on $\mathfrak{H}$. The isotropy group of i is the group of matrices

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \quad \text{with } \theta \text{ real.}$$

This group is usually denoted by K. The group G operates transitively. You can verify all these statements as easy exercises.

Let G operate on a set S. Then two orbits of G are either disjoint or are equal. Indeed, if Gs_1 and Gs_2 are two orbits with an element s in common, then $s = xs_1$ for some $x \in G$, and hence $Gs = Gxs_1 = Gs_1$. Similarly, $Gs = Gs_2$. Hence S is the disjoint union of the distinct orbits, and we can write

$$S = \bigcup_{i \in I} Gs_i \qquad \text{(disjoint)}, \qquad \text{also denoted } S = \coprod_{i \in I} Gs_i,$$

where I is some indexing set, and the s_i are elements of distinct orbits. If S is finite, this gives a decomposition of the order of S as a sum of orders of orbits, which we call the **orbit decomposition formula**, namely

$$\boxed{\operatorname{card}(S) = \sum_{i \in I} (G : G_{s_i}).}$$

Let x, y be elements of a group (or monoid) G. They are said to **commute** if $xy = yx$. If G is a group, the set of all elements $x \in G$ which commute with all elements of G is a subgroup of G which we called the **center** of G. Let G act on itself by conjugation. Then x is in the center if and only if the orbit of x is x itself, and thus has one element. In general, the order of the orbit of x is equal to the index of the normalizer of x. Thus when G is a finite group, the above formula reads

$$\boxed{(G : 1) = \sum_{x \in C} (G : G_x)}$$

where C is a set of representatives for the distinct conjugacy classes, and the sum is taken over all $x \in C$. This formula is also called the **class formula**.

The class formula and the orbit decomposition formula will be used systematically in the next section on Sylow groups, which may be viewed as providing examples for these formulas.

Readers interested in Sylow groups may jump immediately to the next section. The rest of this section deals with special properties of the symmetric group, which may serve as examples of the general notions we have developed.

The symmetric group. Let S_n be the group of permutations of a set with n elements. This set may be taken to be the set of integers $J_n = \{1, 2, \dots, n\}$. Given any $\sigma \in S_n$, and any integer i, $1 \leqq i \leqq n$, we may form the orbit of i under the cyclic group generated by σ. Such an orbit is called a **cycle** for σ, and may be written

$$[i_1 i_2 \cdots i_r], \qquad \text{so} \quad \sigma(i_1) = i_2, \dots, \sigma(i_{r-1}) = i_r,\ \sigma(i_r) = i_1.$$

Then $\{1, \dots, n\}$ may be decomposed into a disjoint union of orbits for the cyclic group generated by σ, and therefore into disjoint cycles. Thus the effect of σ on $\{1, \dots, n\}$ is represented by a product of disjoint cycles.

Example. The cycle [132] represents the permutation σ such that

$$\sigma(1) = 3, \qquad \sigma(3) = 2, \quad \text{and} \quad \sigma(2) = 1.$$

We have $\sigma^2(1) = 2$, $\sigma^3(1) = 1$. Thus $\{1, 3, 2\}$ is the orbit of 1 under the cyclic group generated by σ.

Example. In Exercise 38, one will see how to generate S_n by special types of generators. Perhaps the most important part of that exercise is that if n is prime, σ is an n-cycle and τ is a transposition, then σ, τ generate S_n. As an application in Galois theory, if one tries to prove that a Galois group is all of S_n (as a group of permutations of the roots), it suffices to prove that the Galois group contains an n-cycle and a transposition. See Example 6 of Chapter VI, §2.

We want to associate a sign ± 1 to each permutation. We do this in the standard way. Let f be a function of n variables, say $f: \mathbf{Z}^n \to \mathbf{Z}$, so we can evaluate $f(x_1, \dots, x_n)$. Let σ be a permutation of J_n. We define the function $\pi(\sigma)f$ by

$$\pi(\sigma)f(x_1, \dots, x_n) = f(x_{\sigma(1)}, \dots, x_{\sigma(n)}).$$

Then for $\sigma, \tau \in S_n$ we have $\pi(\sigma\tau) = \pi(\sigma)\pi(\tau)$. Indeed, we use the definition applied to the function $g = \pi(\tau)f$ to get

$$\begin{aligned}\pi(\sigma)\pi(\tau)f(x_1, \dots, x_n) &= (\pi(\tau)f)(x_{\sigma(1)}, \dots, x_{\sigma(n)})\\ &= f(x_{\sigma\tau(1)}, \dots, x_{\sigma\tau(n)})\\ &= \pi(\sigma\tau)f(x_1, \dots, x_n).\end{aligned}$$

Since the identity in S_n operates as the identity on functions, it follows that we have obtained an operation of S_n on the set of functions. We shall write more simply σf instead of $\pi(\sigma)f$. It is immediately verified that for two functions f, g we have

$$\sigma(f + g) = \sigma f + \sigma g \quad \text{and} \quad \sigma(fg) = (\sigma f)(\sigma g).$$

If c is constant, then $\sigma(cf) = c\sigma(f)$.

Proposition 5.3. *There exists a unique homomorphism $\varepsilon\colon S_n \to \{\pm 1\}$ such that for every transposition τ we have $\varepsilon(\tau) = -1$.*

Proof. Let Δ be the function

$$\Delta(x_1, \ldots, x_n) = \prod_{i<j} (x_j - x_i),$$

the product being taken for all pairs of integers i, j satisfying $1 \leqq i < j \leqq n$. Let τ be a transposition, interchanging the two integers r and s. Say $r < s$. We wish to determine

$$\tau\Delta(x_1, \ldots, x_n) = \prod_{i<j} (x_{\tau(j)} - x_{\tau(i)}).$$

For one factor involving $j = s$, $i = r$, we see that τ changes the factor $(x_s - x_r)$ to $-(x_s - x_r)$. All other factors can be considered in pairs as follows:

$$\begin{aligned} &(x_k - x_s)(x_k - x_r) && \text{if } k > s, \\ &(x_s - x_k)(x_k - x_r) && \text{if } r < k < s, \\ &(x_s - x_k)(x_r - x_k) && \text{if } k < r. \end{aligned}$$

Each one of these pairs remains unchanged when we apply τ. Hence we see that $\tau\Delta = -\Delta$.

Let $\varepsilon(\sigma)$ be the sign 1 or -1 such that $\sigma\Delta = \varepsilon(\sigma)\Delta$ for a permutation σ. Since $\pi(\sigma\tau) = \pi(\sigma)\pi(\tau)$, it follows at once that ε is a homomorphism, and the proposition is proved.

In particular, if $\sigma = \tau_1 \cdots \tau_m$ is a product of transpositions, then $\varepsilon(\sigma) = (-1)^m$. As a matter of terminology, we call σ **even** if $\varepsilon(\sigma) = 1$, and **odd** if $\varepsilon(\sigma) = -1$. The even permutations constitute the kernel of ε, which is called the **alternating group** A_n.

Theorem 5.4. *If $n \geqq 5$ then S_n is not solvable.*

Proof. We shall first prove that if H, N are two subgroups of S_n such that $N \subset H$ and N is normal in H, if H contains every 3-cycle, and if H/N is abelian, then N contains every 3-cycle. To see this, let i, j, k, r, s be five distinct integers in J_n, and let $\sigma = [ijk]$ and $\tau = [krs]$. Then a direct computation gives their commutator

$$\sigma\tau\sigma^{-1}\tau^{-1} = [rki].$$

Since the choice of i, j, k, r, s was arbitrary, we see that the cycles $[rki]$ all lie in N for all choices of distinct r, k, i, thereby proving what we wanted.

Now suppose that we have a tower of subgroups

$$S_n = H_0 \supset H_1 \supset H_2 \supset \cdots \supset H_m = \{e\}$$

such that H_v is normal in H_{v-1} for $v = 1, \ldots, m$, and H_v/H_{v-1} is abelian. Since S_n contains every 3-cycle, we conclude that H_1 contains every 3-cycle. By induction, we conclude that $H_m = \{e\}$ contains every 3-cycle, which is impossible, thus proving the theorem.

Remark concerning the sign $\varepsilon(\sigma)$. *A priori*, we defined the sign for a given n, so we should write $\varepsilon_n(\sigma)$. However, suppose $n < m$. Then the restriction of ε_m to S_n (viewed as a permutation of J_n leaving the elements of J_m not in J_n fixed) gives a homomorphism satisfying the conditions of Proposition 5.3, so this restriction is equal to ε_n. Thus $A_m \cap S_n = A_n$.

Next we prove some properties of the alternating group.

(a) *A_n is generated by the 3-cycles. Proof:* Consider the product of two transpositions $[ij][rs]$. If they have an element in common, the product is either the identity or a 3-cycle. If they have no element in common, then

$$[ij][rs] = [ijr][jrs],$$

so the product of two transpositions is also a product of 3-cycles. Since an even permutation is a product of an even number of transpositions, we are done.

(b) *If $n \geqq 5$, all 3-cycles are conjugate in A_n. Proof:* If γ is a permutation, then for a cycle $[i_1 \ldots i_m]$ we have

$$\gamma[i_1 \ldots i_m]\gamma^{-1} = [\gamma(i_1) \ldots \gamma(i_m)].$$

Given 3-cycles $[ijk]$ and $[i'j'k']$ there is a permutation γ such that $\gamma(i) = i'$, $\gamma(j) = j'$, and $\gamma(k) = k'$. Thus two 3-cycles are conjugate in S_n by some element γ. If γ is even, we are done. Otherwise, by assumption $n \geqq 5$ there exist r, s not equal to any one of the three elements i, j, k. Then $[rs]$ commutes with $[ijk]$, and we replace γ by $\gamma[rs]$ to prove (b).

Theorem 5.5. *If $n \geqq 5$ then the alternating group A_n is simple.*

Proof. Let N be a non-trivial normal subgroup of A_n. We prove that N contains some 3-cycle, whence the theorem follows by (b). Let $\sigma \in N$, $\sigma \neq id$, be an element which has the maximal number of fixed points; that is, integers i such that $\sigma(i) = i$. It will suffice to prove that σ is a 3-cycle or the identity. Decompose J_n into disjoint orbits of $\langle\sigma\rangle$.Then some orbits have more than one element. Suppose all orbits have 2 elements (except for the fixed points). Since σ is even, there are at least two such orbits. On their union, σ is represented as a product of two transpositions $[ij][rs]$. Let $k \neq i, j, r, s$. Let $\tau = [rsk]$. Let $\sigma' = \tau\sigma\tau^{-1}\sigma^{-1}$. Then σ' is a product of a conjugate of σ and σ^{-1}, so $\sigma' \in N$. But σ' leaves i, j fixed, and any element $t \in J_n$, $t \neq i, j, r, s, k$ left fixed by σ is also fixed by σ', so σ' has more fixed points than σ, contradicting our hypothesis.

So we are reduced to the case when at least one orbit of $\langle\sigma\rangle$ has $\geqq 3$ elements, say $i, j, k, \ldots$. If σ is not the 3-cycle $[ijk]$, then σ must move at least two other elements of J_n, otherwise σ is an odd permutation $[ijkr]$ for some $r \in J_n$, which is impossible. Then let σ move r, s other than i, j, k, and let $\tau = [krs]$. Let σ' be the commutator as before. Then $\sigma' \in N$ and $\sigma'(i) = i$, and all fixed points of σ are also fixed points of σ' whence σ' has more fixed points than σ, a contradiction which proves the theorem.

Example. For $n = 4$, the group A_4 is not simple. As an exercise, show that A_4 contains a unique subgroup of order 4, which is not cyclic, and which is normal. This subgroup is also normal in S_4. Write down explicitly its elements as products of transpositions.

§6. SYLOW SUBGROUPS

Let p be a prime number. By a **p-group**, we mean a finite group whose order is a power of p (i.e. p^n for some integer $n \geqq 0$). Let G be a finite group and H a subgroup. We call H a **p-subgroup** of G if H is a p-group. We call H a **p-Sylow** subgroup if the order of H is p^n and if p^n is the highest power of p dividing the order of G. We shall prove below that such subgroups always exist. For this we need a lemma.

Lemma 6.1. *Let G be a finite abelian group of order m, let p be a prime number dividing m. Then G has a subgroup of order p.*

Proof. We first prove by induction that if G has exponent n then the order of G divides some power of n. Let $b \in G$, $b \neq 1$, and let H be the cyclic subgroup generated by b. Then the order of H divides n since $b^n = 1$, and n is an exponent for G/H. Hence the order of G/H divides a power of n by induction, and consequently so does the order of G because

$$(G:1) = (G:H)(H:1).$$

Let G have order divisible by p. By what we have just seen, there exists an element x in G whose period is divisible by p. Let this period be ps for some integer s. Then $x^s \neq 1$ and obviously x^s has period p, and generates a subgroup of order p, as was to be shown.

Theorem 6.2. *Let G be a finite group and p a prime number dividing the order of G. Then there exists a p-Sylow subgroup of G.*

Proof. By induction on the order of G. If the order of G is prime, our assertion is obvious. We now assume given a finite group G, and assume the theorem proved for all groups of order smaller than that of G. If there exists a proper subgroup H of G whose index is prime to p, then a p-Sylow subgroup of H will also be one of G, and our assertion follows by induction. We may therefore assume that every proper subgroup has an index divisible by p. We now let G act on itself by conjugation. From the class formula we obtain

$$(G:1) = (Z:1) + \sum (G:G_x).$$

Here, Z is the center of G, and the term $(Z:1)$ corresponds to the orbits having one element, namely the elements of Z. The sum on the right is taken over the other orbits, and each index $(G:G_x)$ is then > 1, hence divisible by p. Since p divides the order of G, it follows that p divides the order of Z, hence in particular that G has a non-trivial center.

Let a be an element of order p in Z, and let H be the cyclic group generated by a. Since H is contained in Z, it is normal. Let $f: G \to G/H$ be the canonical map. Let p^n be the highest power of p dividing $(G:1)$. Then p^{n-1} divides the order of G/H. Let K' be a p-Sylow subgroup of G/H (by induction) and let $K = f^{-1}(K')$. Then $K \supset H$ and f maps K onto K'. Hence we have an isomorphism $K/H \approx K'$. Hence K has order $p^{n-1}p = p^n$, as desired.

For the rest of the theorems, we systematically use the notion of a fixed point. Let G be a group operating on a set S. Recall that a **fixed point** s of G in S is an element s of S such that $xs = s$ for all $x \in G$.

Lemma 6.3. *Let H be a p-group acting on a finite set S. Then:*

(a) *The number of fixed points of H is $\equiv \#(S) \bmod p$.*

(b) *If H has exactly one fixed point, then $\#(S) \equiv 1 \bmod p$.*

(c) *If $p \mid \#(S)$, then the number of fixed points of H is $\equiv 0 \bmod p$.*

Proof. We repeatedly use the orbit formula

$$\#(S) = \sum (H : H_{s_i}).$$

For each fixed point s_i we have $H_{s_i} = H$. For s_i not fixed, the index $(H : H_{s_i})$ is divisible by p, so (a) follows at once. Parts (b) and (c) are special cases of (a), thus proving the lemma.

Remark. In Lemma 6.3(c), if H has one fixed point, then H has at least p fixed points.

Theorem 6.4. *Let G be a finite group.*

(i) *If H is a p-subgroup of G, then H is contained in some p-Sylow subgroup.*

(ii) *All p-Sylow subgroups are conjugate.*

(iii) *The number of p-Sylow subgroups of G is* $\equiv 1 \bmod p$.

Proof. Let P be a p-Sylow subgroup of G. Suppose first that H is contained in the normalizer of P. We prove that $H \subset P$. Indeed, HP is then a subgroup of the normalizer, and P is normal in HP. But

$$(HP : P) = (H : H \cap P),$$

so if $HP \neq P$, then HP has order a power of p, and the order is larger than $\#(P)$, contradicting the hypothesis that P is a Sylow group. Hence $HP = P$ and $H \subset P$.

Next, let S be the set of all conjugates of P in G. Then G operates on S by conjugation. Since the normalizer of P contains P, and has therefore index prime to p, it follows that $\#(S)$ is not divisible by p. Now let H be any p-subgroup. Then H also acts on S by conjugation. By Lemma 6.3(a), we know that H cannot have 0 fixed points. Let Q be a fixed point. By definition this means that H is contained in the normalizer of Q, and hence by the first part of the proof, that $H \subset Q$, which proves the first part of the theorem. The second part follows immediately by taking H to be a p-Sylow group, so $\#(H) = \#(Q)$, whence $H = Q$. In particular, when H is a p-Sylow group, we see that H has only one fixed point, so that **(iii)** follows from Lemma 6.3(b). This proves the theorem.

Theorem 6.5. *Let G be a finite p-group. Then G is solvable. If its order is* > 1, *then G has a non-trivial center.*

Proof. The first assertion follows from the second, since if G has center Z, and we have an abelian tower for G/Z by induction, we can lift this abelian tower to G to show that G is solvable. To prove the second assertion, we use the class equation

$$(G : 1) = \text{card}(Z) + \sum (G : G_x),$$

the sum being taken over certain x for which $(G : G_x) \neq 1$. Then p divides $(G : 1)$ and also divides every term in the sum, so that p divides the order of the center, as was to be shown.

Corollary 6.6. *Let G be a p-group which is not of order* 1. *Then there exists a sequence of subgroups*

$$\{e\} = G_0 \subset G_1 \subset G_2 \subset \cdots \subset G_n = G$$

such that G_i is normal in G and G_{i+1}/G_i is cyclic of order p.

Proof. Since G has a non-trivial center, there exists an element $a \neq e$ in the center of G, and such that a has order p. Let H be the cyclic group generated by a. By induction, if $G \neq H$, we can find a sequence of subgroups as stated above in the factor group G/H. Taking the inverse image of this tower in G gives us the desired sequence in G.

We now give some examples to show how to put some of the group theory together.

Lemma 6.7. *Let G be a finite group and let p be the smallest prime dividing the order of G. Let H be a subgroup of index p. Then H is normal.*

Proof. Let $N(H) = N$ be the normalizer of H. Then $N = G$ or $N = H$. If $N = G$ we are done. Suppose $N = H$. Then the orbit of H under conjugation has $p = (G : H)$ elements, and the representation of G on this orbit gives a homomorphism of G into the symmetric group on p elements, whose order is $p!$. Let K be the kernel. Then K is the intersection of the isotropy groups, and the isotropy group of H is H by assumption, so $K \subset H$. If $K \neq H$, then from

$$(G : K) = (G : H)(H : K) = p(H : K),$$

and the fact that only the first power of p divides $p!$, we conclude that some prime dividing $(p - 1)!$ also divides $(H : K)$, which contradicts the assumption that p is the smallest prime dividing the order of G, and proves the lemma.

Proposition 6.8. *Let p, q be distinct primes and let G be a group of order pq. Then G is solvable.*

Proof. Say $p < q$. Let Q be a Sylow subgroup of order q. Then Q has index p, so by the lemma, Q is normal and the factor group has order p. But a group of prime order is cyclic, whence the proposition follows.

Example. Let G be a group of order 35. We claim that G is cyclic.

Proof. Let H_7 be the Sylow subgroup of order 7. Then H_7 is normal by Lemma 6.7. Let H_5 be a 5-Sylow subgroup, which is of order 5. Then H_5 operates by conjugation on H_7, so we get a homomorphism $H_5 \to \mathrm{Aut}(H_7)$. But $\mathrm{Aut}(H_7)$ is cyclic of order 6, so $H_5 \to \mathrm{Aut}(H_7)$ is trivial, so every element of H_5 commutes with elements of H_7. Let $H_5 = \langle x \rangle$ and $H_7 = \langle y \rangle$. Then x, y commute with each other and with themselves, so G is abelian, and so G is cyclic by Proposition 4.3(**v**).

Example. The techniques which have been developed are sufficient to treat many cases of the above types. For instance every group of order < 60 is solvable, as you will prove in Exercise 27.

§7. DIRECT SUMS AND FREE ABELIAN GROUPS

Let $\{A_i\}_{i \in I}$ be a family of abelian groups. We define their **direct sum**

$$A = \bigoplus_{i \in I} A_i$$

to be the subset of the direct product $\prod A_i$ consisting of all families $(x_i)_{i \in I}$ with

$x_i \in A_i$ such that $x_i = 0$ for all but a finite number of indices i. Then it is clear that A is a subgroup of the product. For each index $j \in I$, we map

$$\lambda_j : A_j \to A$$

by letting $\lambda_j(x)$ be the element whose j-th component is x, and having all other components equal to 0. Then λ_i is an injective homomorphism.

Proposition 7.1. *Let $\{f_i : A_i \to B\}$ be a family of homomorphisms into an abelian group B. Let $A = \bigoplus A_i$. There exists a unique homomorphism*

$$f : A \to B$$

such that $f \circ \lambda_j = f_j$ for all j.

Proof. We can define a map $f: A \to B$ by the rule

$$f((x_i)_{i \in I}) = \sum_{i \in I} f_i(x_i).$$

The sum on the right is actually finite since all but a finite number of terms are 0. It is immediately verified that our map f is a homomorphism. Furthermore, we clearly have $f \circ \lambda_j(x) = f_j(x)$ for each j and each $x \in A_j$. Thus f has the desired commutativity property. It is also clear that the map f is uniquely determined, as was to be shown.

The property expressed in Proposition 7.1 is called the **universal property** of the direct sum. Cf. §11.

Example. Let A be an abelian group, and let $\{A_i\}_{i \in I}$ be a family of subgroups. Then we get a homomorphism

$$\bigoplus_{i \in I} A_i \to A \quad \text{such that} \quad (x_i) \mapsto \sum x_i.$$

Theorem 8.1 will provide an important specific application.

Let A be an abelian group and B, C subgroups. If $B + C = A$ and $B \cap C = \{0\}$ then the map

$$B \times C \to A$$

given by $(x, y) \mapsto x + y$ is an isomorphism (as we already noted in the non-commutative case). Instead of writing $A = B \times C$ we shall write

$$A = B \oplus C$$

and say that A is the **direct sum** of B and C. We use a similar notation for the direct sum of a finite number of subgroups $B_1, \ldots, B_n$ such that

$$B_1 + \cdots + B_n = A$$

and

$$B_{i+1} \cap (B_1 + \cdots + B_i) = 0.$$

In that case we write

$$A = B_1 \oplus \cdots \oplus B_n.$$

Let A be an abelian group. Let $\{e_i\}$ $(i \in I)$ be a family of elements of A. We say that this family is a **basis** for A if the family is not empty, and if every element of A has a unique expression as a linear combination

$$x = \sum x_i e_i$$

with $x_i \in \mathbf{Z}$ and almost all $x_i = 0$. Thus the sum is actually a finite sum. An abelian group is said to be **free** if it has a basis. If that is the case, it is immediate that if we let $Z_i = \mathbf{Z}$ for all i, then A is isomorphic to the direct sum

$$A \approx \bigoplus_{i \in I} Z_i.$$

Next let S be a set. We shall define the free abelian group generated by S as follows. Let $\mathbf{Z}\langle S\rangle$ be the set of all maps $\varphi : S \to \mathbf{Z}$ such that $\varphi(x) = 0$ for almost all $x \in S$. Then $\mathbf{Z}\langle S\rangle$ is an abelian group (addition being the usual addition of maps). If k is an integer and x is an element of S, we denote by $k \cdot x$ the map φ such that $\varphi(x) = k$ and $\varphi(y) = 0$ if $y \neq x$. Then it is obvious that every element φ of $\mathbf{Z}\langle S\rangle$ can be written in the form

$$\varphi = k_1 \cdot x_1 + \cdots + k_n \cdot x_n$$

for some integers k_i and elements $x_i \in S$ $(i = 1, \ldots, n)$, all the x_i being distinct. Furthermore, φ *admits a unique such expression*, because if we have

$$\varphi = \sum_{x \in S} k_x \cdot x = \sum_{x \in S} k'_x \cdot x$$

then

$$0 = \sum_{x \in S} (k_x - k'_x) \cdot x,$$

whence $k'_x = k_x$ for all $x \in S$.

We map S into $\mathbf{Z}\langle S\rangle$ by the map $f_S = f$ such that $f(x) = 1 \cdot x$. It is then clear that f is injective, and that $f(S)$ generates $\mathbf{Z}\langle S\rangle$. If $g : S \to B$ is a mapping of S into some abelian group B, then we can define a map

$$g_* : \mathbf{Z}\langle S\rangle \to B$$

such that

$$g_*\left(\sum_{x \in S} k_x \cdot x\right) = \sum_{x \in S} k_x g(x).$$

This map is a homomorphism (trivial) and we have $g_* \circ f = g$ (also trivial). It is the only homomorphism which has this property, for any such homomorphism g_* must be such that $g_*(1 \cdot x) = g(x)$.

It is customary to identify S in $\mathbf{Z}\langle S\rangle$, and we sometimes omit the dot when we write $k_x x$ or a sum $\sum k_x x$.

If $\lambda : S \to S'$ is a mapping of sets, there is a unique homomorphism $\bar{\lambda}$ making the following diagram commutative:

$$\begin{tikzcd}
S \arrow[r, "f_S"] \arrow[d, "\lambda"'] & \mathbf{Z}\langle S\rangle \arrow[d, "\bar{\lambda}"] \\
S' \arrow[r, "f_{S'}"'] & \mathbf{Z}\langle S'\rangle
\end{tikzcd}$$

In fact, $\bar{\lambda}$ is none other than $(f_{S'} \circ \lambda)_*$, with the notation of the preceding paragraph. The proof of this statement is left as a trivial exercise.

We shall denote $\mathbf{Z}\langle S\rangle$ also by $F_{\text{ab}}(S)$, and call $F_{\text{ab}}(S)$ the **free abelian group generated by** S. We call elements of S its **free generators**.

As an exercise, show that every abelian group A is isomorphic to a factor group of a free abelian group F. If A is finitely generated, show that one can select F to be finitely generated also.

If the set S above consists of n elements, then we say that the free abelian group $F_{\text{ab}}(S)$ is the free abelian group on n generators. If S is the set of n letters $x_1, \dots, x_n$, we say that $F_{\text{ab}}(S)$ is the free abelian group with free generators $x_1, \dots, x_n$.

An abelian group is **free** if and only if it is isomorphic to a free abelian group $F_{\text{ab}}(S)$ for some set S. Let A be an abelian group, and let S be a basis for A. Then it is clear that A is isomorphic to the free abelian group $F_{\text{ab}}(S)$.

As a matter of notation, if A is an abelian group and T a subset of elements of A, we denote by $\langle T\rangle$ the subgroup generated by the elements of T, i.e., the smallest subgroup of A containing T.

Example. The Grothendieck group. Let M be a commutative monoid, written additively. There exists a commutative group $K(M)$ and a monoid-homomorphism

$$\gamma : M \to K(M)$$

having the following universal property. If $f : M \to A$ is a homomorphism into an abelian group A, then there exists a unique homomorphism $f_*: K(M) \to A$ making the following diagram commutative:

$$\begin{tikzcd}
M \arrow[rr, "\gamma"] \arrow[dr, "f"'] & & K(M) \arrow[dl, "f_*"] \\
& A &
\end{tikzcd}$$

Proof. Let $F_{\text{ab}}(M)$ be the free abelian group generated by M. We denote the generator of $F_{\text{ab}}(M)$ corresponding to an element $x \in M$ by $[x]$. Let B be the subgroup generated by all elements of type

$$[x + y] - [x] - [y]$$

where $x, y \in M$. We let $K(M) = F_{\text{ab}}(M)/B$, and let

$$\gamma : M \to K(M)$$

be the map obtained by composing the injection of M into $F_{\text{ab}}(M)$ given by $x \mapsto [x]$, and the canonical map

$$F_{\text{ab}}(M) \to F_{\text{ab}}(M)/B.$$

It is then clear that γ is a homomorphism, and satisfies the desired universal property.

The universal group $K(M)$ is called the **Grothendieck group**.

We shall say that the **cancellation law** holds in M if, whenever $x, y, z \in M$, and $x + z = y + z$, we have $x = y$.

We then have an important criterion when the universal map γ above is injective:

If the cancellation law holds in M, then the canonical map γ of M into its Grothendieck group is injective.

Proof. This is essentially the same proof as when one constructs the integers from the natural numbers. We consider pairs (x, y) with $x, y \in M$ and say that (x, y) is equivalent to (x', y') if $y + x' = x + y'$. We define addition of pairs componentwise. Then the equivalence classes of pairs form a group, whose 0 element is the class of $(0, 0)$ [or the class of (x, x) for any $x \in M$]. The negative of an element (x, y) is (y, x). We have a homomorphism

$$x \mapsto \text{class of } (0, x)$$

which is injective, as one sees immediately by applying the cancellation law. Thus we have constructed a homomorphism of M into a group, which is injective. It follows that the universal homomorphism must also be injective.

Examples. See the example of projective modules in Chapter III, §4. For a relatively fancy context, see: K. Kato, Logarithmic structures of Fontaine-Illusie, *Algebraic Geometry, Analysis and Number Theory, Proc. JAMI Conference*, J. Igusa (Ed.), Johns Hopkins Press (1989) pp. 195–224.

Given an abelian group A and a subgroup B, it is sometimes desirable to find a subgroup C such that $A = B \oplus C$. The next lemma gives us a condition under which this is true.

Lemma 7.2. *Let $A \xrightarrow{f} A'$ be a surjective homomorphism of abelian groups, and assume that A' is free. Let B be the kernel of f. Then there exists a subgroup C of A such that the restriction of f to C induces an isomorphism of C with A', and such that $A = B \oplus C$.*

Proof. Let $\{x_i'\}_{i \in I}$ be a basis of A', and for each $i \in I$, let x_i be an element of A such that $f(x_i) = x_i'$. Let C be the subgroup of A generated by all elements x_i, $i \in I$. If we have a relation

$$\sum_{i \in I} n_i x_i = 0$$

with integers n_i, almost all of which are equal to 0, then applying f yields

$$0 = \sum_{i \in I} n_i f(x_i) = \sum_{i \in I} n_i x'_i,$$

whence all $n_i = 0$. Hence our family $\{x_i\}_{i \in I}$ is a basis of C. Similarly, one sees that if $z \in C$ and $f(z) = 0$ then $z = 0$. Hence $B \cap C = 0$. Let $x \in A$. Since $f(x) \in A'$ there exist integers n_i, $i \in I$, such that

$$f(x) = \sum_{i \in I} n_i x'_i.$$

Applying f to $x - \sum_{i \in I} n_i x_i$, we find that this element lies in the kernel of f, say

$$x - \sum_{i \in I} n_i x_i = b \in B.$$

From this we see that $x \in B + C$, and hence finally that $A = B \oplus C$ is a direct sum, as contended.

Theorem 7.3. *Let A be a free abelian group, and let B be a subgroup. Then B is also a free abelian group, and the cardinality of a basis of B is $\leqq$ the cardinality of a basis for A. Any two bases of B have the same cardinality.*

Proof. We shall give the proof only when A is finitely generated, say by a basis $\{x_1, \ldots, x_n\}$ $(n \geqq 1)$, and give the proof by induction on n. We have an expression of A as direct sum:

$$A = \mathbf{Z}x_1 \oplus \cdots \oplus \mathbf{Z}x_n.$$

Let $f: A \to \mathbf{Z}x_1$ be the projection, i.e. the homomorphism such that

$$f(m_1 x_1 + \cdots + m_n x_n) = m_1 x_1$$

whenever $m_i \in \mathbf{Z}$. Let B_1 be the kernel of $f|B$. Then B_1 is contained in the free subgroup $\langle x_2, \ldots, x_n \rangle$. By induction, B_1 is free and has a basis with $\leqq n - 1$ elements. By the lemma, there exists a subgroup C_1 isomorphic to a subgroup of $\mathbf{Z}x_1$ (namely the image of $f|B$) such that

$$B = B_1 \oplus C_1.$$

Since $f(B)$ is either 0 or infinite cyclic, i.e. free on one generator, this proves that B is free.

(When A is not finitely generated, one can use a similar transfinite argument. See Appendix 2, §2, the example after Zorn's Lemma.)

We also observe that our proof shows that there exists at least one basis of B whose cardinality is $\leqq n$. We shall therefore be finished when we prove the last statement, that any two bases of B have the same cardinality. Let S be one basis, with a finite number of elements m. Let T be another basis, and suppose that T has at least r elements. It will suffice to prove that $r \leqq m$ (one can then use symmetry). Let p be a prime number. Then B/pB is a direct sum of cyclic groups of order p, with m terms in the sum. Hence its order is p^m. Using the basis T instead of S, we conclude that B/pB contains an r-fold product of cyclic groups of order p, whence $p^r \leqq p^m$, and $r \leqq m$, as was to be shown. (Note that we did not assume a priori that T was finite.)

The number of elements in a basis of a free abelian group A will be called the **rank** of A.

§8. FINITELY GENERATED ABELIAN GROUPS

The groups referred to in the title of this section occur so frequently that it is worth while to state a theorem which describes their structure completely. Throughout this section we write our abelian groups additively.

Let A be an abelian group. An element $a \in A$ is said to be a **torsion** element if it has finite period. The subset of all torsion elements of A is a subgroup of A called the **torsion subgroup** of A. (If a has period m and b has period n then, writing the group law additively, we see that $a \pm b$ has a period dividing mn.)

The torsion subgroup of A is denoted by A_{tor}, or simply A_t. An abelian group is called a **torsion group** if $A = A_{\text{tor}}$, that is all elements of A are of finite order.

A finitely generated torsion abelian group is obviously finite. We shall begin by studying torsion abelian groups. If A is an abelian group and p a prime number, we denote by $A(p)$ the subgroup of all elements $x \in A$ whose period is a power of p. Then $A(p)$ is a torsion group, and is a p-group if it is finite.

Theorem 8.1 *Let A be a torsion abelian group. Then A is the direct sum of its subgroups $A(p)$ for all primes p such that $A(p) \neq 0$.*

Proof. There is a homomorphism

$$\bigoplus_p A(p) \to A$$

which to each element (x_p) in the direct sum associates the element $\sum x_p$ in A. We prove that this homomorphism is both surjective and injective. Suppose x is in the kernel, so $\sum x_p = 0$. Let q be a prime. Then

$$x_q = \sum_{p \neq q} (-x_p).$$

Let m be the least common multiple of the periods of elements x_p on the right-hand side, with $x_q \neq 0$ and $p \neq q$. Then $mx_q = 0$. But also $q^r x_q = 0$ for some positive integer r. If d is the greatest common divisor of m, q^r then $dx_q = 0$, but $d = 1$, so $x_q = 0$. Hence the kernel is trivial, and the homomorphism is injective.

As for the surjectivity, for each positive integer m, denote by A_m the kernel of multiplication by m, i.e. the subgroup of $x \in A$ such that $mx = 0$. We prove:

If $m = rs$ with r, s positive relative prime integers, then $A_m = A_r + A_s$.

Indeed, there exist integers u, v such that $ur + vs = 1$. Then $x = urx + vsx$, and $urx \in A_s$ while $vsx \in A_r$, and our assertion is proved. Repeating this process inductively, we conclude:

$$\text{If } m = \prod_{p|m} p^{e(p)} \text{ then } A_m = \sum_{p|m} A_{p^{e(p)}}.$$

Hence the map $\bigoplus A(p) \to A$ is surjective, and the theorem is proved.

Example. Let $A = \mathbf{Q}/\mathbf{Z}$. Then $\mathbf{Q}/\mathbf{Z}$ is a torsion abelian group, isomorphic to the direct sum of its subgroups $(\mathbf{Q}/\mathbf{Z})(p)$. Each $(\mathbf{Q}/\mathbf{Z})(p)$ consists of those elements which can be represented by a rational number a/p^k with $a \in \mathbf{Z}$ and k some positive integer, i.e. a rational number having only a p-power in the denominator. See also Chapter IV, Theorem 5.1.

In what follows we shall deal with finite abelian groups, so only a finite number of primes (dividing the order of the group) will come into play. In this case, the direct sum is "the same as" the direct product.

Our next task is to describe the structure of finite abelian p-groups. Let $r_1, \ldots, r_s$ be integers $\geqq 1$. A finite p-group A is said to be of **type** $(p^{r_1}, \ldots, p^{r_s})$ if A is isomorphic to the product of cyclic groups of orders p^{r_i} $(i = 1, \ldots, s)$. We shall need the following remark.

Remark. Let A be a finite abelian p-group. Let b be an element of A, $b \neq 0$. Let k be an integer $\geqq 0$ such that $p^k b \neq 0$, and let p^m be the period of $p^k b$. Then b has period p^{k+m}. [*Proof*: We certainly have $p^{k+m}b = 0$, and if $p^n b = 0$ then first $n \geqq k$, and second $n \geqq k + m$, otherwise the period of $p^k b$ would be smaller than p^m.]

Theorem 8.2. *Every finite abelian p-group is isomorphic to a product of cyclic p-groups. If it is of type $(p^{r_1}, \ldots, p^{r_s})$ with*

$$r_1 \geqq r_2 \geqq \cdots \geqq r_s \geqq 1,$$

then the sequence of integers $(r_1, \ldots, r_s)$ is uniquely determined.

Proof. We shall prove the existence of the desired product by induction. Let $a_1 \in A$ be an element of maximal period. We may assume without loss of generality that A is not cyclic. Let A_1 be the cyclic subgroup generated by a_1, say of period p^{r_1}. We need a lemma.

Lemma 8.3. *Let $\bar{b}$ be an element of A/A_1, of period p^r. Then there exists a representative a of $\bar{b}$ in A which also has period p^r.*

Proof. Let b be any representative of $\bar{b}$ in A. Then $p^r b$ lies in A_1, say $p^r b = na_1$ with some integer $n \geqq 0$. We note that the period of $\bar{b}$ is $\leqq$ the period of b. If $n = 0$ we are done. Otherwise write $n = p^k \mu$ where μ is prime to p. Then μa_1 is also a generator of A_1, and hence has period p^{r_1}. We may assume $k \leqq r_1$. Then $p^k \mu a_1$ has period $p^{r_1 - k}$. By our previous remarks, the element b has period

$$p^{r + r_1 - k}$$

whence by hypothesis, $r + r_1 - k \leqq r_1$ and $r \leqq k$. This proves that there exists an element $c \in A_1$ such that $p^r b = p^r c$. Let $a = b - c$. Then a is a representative for $\bar{b}$ in A and $p^r a = 0$. Since period $(a) \leqq p^r$ we conclude that a has period equal to p^r.

We return to the main proof. By induction, the factor group A/A_1 has a product expression

$$A/A_1 = \bar{A}_2 \times \cdots \times \bar{A}_s$$

into cyclic subgroups of orders $p^{r_2}, \ldots, p^{r_s}$ respectively, and we may assume $r_2 \geqq \cdots \geqq r_s$. Let $\bar{a}_i$ be a generator for $\bar{A}_i$ $(i = 2, \ldots, s)$ and let a_i be a representative in A of the same period as $\bar{a}_i$. Let A_i be the cyclic subgroup generated by a_i. We contend that A is the direct sum of $A_1, \ldots, A_s$.

Given $x \in A$, let $\bar{x}$ denote its residue class in A/A_1. There exist integers $m_i \geqq 0$ $(i = 2, \ldots, s)$ such that

$$\bar{x} = m_2 \bar{a}_2 + \cdots + m_s \bar{a}_s.$$

Hence $x - m_2 a_2 - \cdots - m_s a_s$ lies in A_1, and there exists an integer $m_1 \geqq 0$ such that

$$x = m_1 a_1 + m_2 a_2 + \cdots + m_s a_s.$$

Hence $A_1 + \cdots + A_s = A$.

Conversely, suppose that $m_1, \ldots, m_s$ are integers $\geqq 0$ such that

$$0 = m_1 a_1 + \cdots + m_s a_s.$$

Since a_i has period p^{r_i} $(i = 1, \ldots, s)$, we may suppose that $m_i < p^{r_i}$. Putting a bar on this equation yields

$$0 = m_2 \bar{a}_2 + \cdots + m_s \bar{a}_s.$$

Since A/A_1 is a direct product of $\bar{A}_2, \ldots, \bar{A}_s$ we conclude that each $m_i = 0$ for $i = 2, \ldots, s$. But then $m_1 = 0$ also, and hence all $m_i = 0$ $(i = 1, \ldots, s)$. From this it follows at once that

$$(A_1 + \cdots + A_i) \cap A_{i+1} = 0$$

for each $i \geqq 1$, and hence that A is the direct product of $A_1, \ldots, A_s$, as desired.

We prove uniqueness, by induction. Suppose that A is written in two ways as a direct sum of cyclic groups, say of type

$$(p^{r_1}, \ldots, p^{r_s}) \quad \text{and} \quad (p^{m_1}, \ldots, p^{m_k})$$

with $r_1 \geqq \cdots \geqq r_s \geqq 1$ and $m_1 \geqq \cdots \geqq m_k \geqq 1$. Then pA is also a p-group, of order strictly less than the order of A, and is of type

$$(p^{r_1-1}, \ldots, p^{r_s-1}) \quad \text{and} \quad (p^{m_1-1}, \ldots, p^{m_k-1}),$$

it being understood that if some exponent r_i or m_j is equal to 1, then the factor corresponding to

$$p^{r_i-1} \quad \text{or} \quad p^{m_j-1}$$

in pA is simply the trivial group 0. By induction, the subsequence of

$$(r_1 - 1, \ldots, r_s - 1)$$

consisting of those integers $\geqq 1$ is uniquely determined, and is the same as the corresponding subsequence of

$$(m_1 - 1, \ldots, m_k - 1).$$

In other words, we have $r_i - 1 = m_i - 1$ for all those integers i such that $r_i - 1$ or $m_i - 1 \geqq 1$. Hence $r_i = m_i$ for all these integers i, and the two sequences

$$(p^{r_1}, \ldots, p^{r_s}) \quad \text{and} \quad (p^{m_1}, \ldots, p^{m_k})$$

can differ only in their last components which can be equal to p. These correspond to factors of type $(p, \ldots, p)$ occurring say ν times in the first sequences and μ times in the second sequence. Thus for some integer n, A is of type

$$(p^{r_1}, \ldots, p^{r_n}, \underbrace{p, \ldots, p}_{\nu \text{ times}}) \quad \text{and} \quad (p^{r_1}, \ldots, p^{r_n}, \underbrace{p, \ldots, p}_{\mu \text{ times}}).$$

Thus the order of A is equal to

$$p^{r_1 + \cdots + r_n} p^{\nu} = p^{r_1 + \cdots + r_n} p^{\mu},$$

whence $\nu = \mu$, and our theorem is proved.

A group G is said to be **torsion free**, or without torsion, if whenever an element x of G has finite period, then x is the unit element.

Theorem 8.4. *Let A be a finitely generated torsion-free abelian group. Then A is free.*

Proof. Assume $A \neq 0$. Let S be a finite set of generators, and let $x_1, \ldots, x_n$ be a maximal subset of S having the property that whenever $\nu_1, \ldots, \nu_n$ are integers such that

$$\nu_1 x_1 + \cdots + \nu_n x_n = 0,$$

then $\nu_j = 0$ for all j. (Note that $n \geqq 1$ since $A \neq 0$). Let B be the subgroup generated by $x_1, \ldots, x_n$. Then B is free. Given $y \in S$ there exist integers $m_1, \ldots, m_n, m$ not all zero such that

$$my + m_1 x_1 + \cdots + m_n x_n = 0,$$

by the assumption of maximality on $x_1, \dots, x_n$. Furthermore, $m \neq 0$; otherwise all $m_j = 0$. Hence my lies in B. This is true for every one of a finite set of generators y of A, whence there exists an integer $m \neq 0$ such that $mA \subset B$. The map

$$x \mapsto mx$$

of A into itself is a homomorphism, having trivial kernel since A is torsion free. Hence it is an isomorphism of A onto a subgroup of B. By Theorem 7.3 of the preceding section, we conclude that mA is free, whence A is free.

Theorem 8.5. *Let A be a finitely generated abelian group, and let A_{tor} be the subgroup consisting of all elements of A having finite period. Then A_{tor} is finite, and A/A_{tor} is free. There exists a free subgroup B of A such that A is the direct sum of A_{tor} and B.*

Proof. We recall that a finitely generated torsion abelian group is obviously finite. Let A be finitely generated by n elements, and let F be the free abelian group on n generators. By the universal property, there exists a surjective homomorphism

$$F \xrightarrow{\varphi} A$$

of F onto A. The subgroup $\varphi^{-1}(A_{\text{tor}})$ of F is finitely generated by Theorem 7.3. Hence A_{tor} itself is finitely generated, hence finite.

Next, we prove that A/A_{tor} has no torsion. Let $\bar{x}$ be an element of A/A_{tor} such that $m\bar{x} = 0$ for some integer $m \neq 0$. Then for any representative of x of $\bar{x}$ in A, we have $mx \in A_{\text{tor}}$, whence $qmx = 0$ for some integer $q \neq 0$. Then $x \in A_{\text{tor}}$, so $\bar{x} = 0$, and A/A_{tor} is torsion free. By Theorem 8.4, A/A_{tor} is free. We now use the lemma of Theorem 7.3 to conclude the proof.

The rank of A/A_{tor} is also called the **rank** of A.

For other contexts concerning Theorem 8.5, see the structure theorem for modules over principal rings in Chapter III, §7, and Exercises 5, 6, and 7 of Chapter III.

§9. THE DUAL GROUP

Let A be an abelian group of exponent $m \geqq 1$. This means that for each element $x \in A$ we have $mx = 0$. Let Z_m be a cyclic group of order m. We denote by $A^\wedge$, or $\text{Hom}(A, Z_m)$ the group of homomorphisms of A into Z_m, and call it the **dual** of A.

Let $f : A \to B$ be a homomorphism of abelian groups, and assume both have exponent m. Then f induces a homomorphism

$$f^\wedge : B^\wedge \to A^\wedge.$$

Namely, for each $\psi \in B^\wedge$ we define $f^\wedge(\psi) = \psi \circ f$. It is trivially verified that $f^\wedge$ is a homomorphism. The properties

$$\mathrm{id}^\wedge = \mathrm{id} \quad \text{and} \quad (f \circ g)^\wedge = g^\wedge \circ f^\wedge$$

are trivially verified.

Theorem 9.1. *If A is a finite abelian group, expressed as a product $A = B \times C$, then $A^\wedge$ is isomorphic to $B^\wedge \times C^\wedge$ (under the mapping described below). A finite abelian group is isomorphic to its own dual.*

Proof. Consider the two projections

$$\begin{tikzcd}
& B \times C \arrow[dl, "f"'] \arrow[dr, "g"] & \\
B & & C
\end{tikzcd}$$

of $B \times C$ on its two components. We get homomorphisms

$$\begin{tikzcd}
& (B \times C)^\wedge & \\
B^\wedge \arrow[ur, "f^\wedge"] & & C^\wedge \arrow[ul, "g^\wedge"']
\end{tikzcd}$$

and we contend that these homomorphisms induce an isomorphism of $B^\wedge \times C^\wedge$ onto $(B \times C)^\wedge$.

In fact, let ψ_1, ψ_2 be in $\mathrm{Hom}(B, Z_m)$ and $\mathrm{Hom}(C, Z_m)$ respectively. Then $(\psi_1, \psi_2) \in B^\wedge \times C^\wedge$, and we have a corresponding element of $(B \times C)^\wedge$ by defining

$$(\psi_1, \psi_2)(x, y) = \psi_1(x) + \psi_2(y),$$

for $(x, y) \in B \times C$. In this way we get a homomorphism

$$B^\wedge \times C^\wedge \to (B \times C)^\wedge.$$

Conversely, let $\psi \in (B \times C)^\wedge$. Then

$$\psi(x, y) = \psi(x, 0) + \psi(0, y).$$

The function ψ_1 on B such that $\psi_1(x) = \psi(x, 0)$ is in $B^\wedge$, and similarly the function ψ_2 on C such that $\psi_2(y) = \psi(0, y)$ is in $C^\wedge$. Thus we get a homomorphism

$$(B \times C)^\wedge \to B^\wedge \times C^\wedge,$$

which is obviously inverse to the one we defined previously. Hence we obtain an isomorphism, thereby proving the first assertion in our theorem.

We can write any finite abelian group as a product of cyclic groups. Thus to prove the second assertion, it will suffice to deal with a cyclic group.

Let A be cyclic, generated by one element x of period n. Then $n \mid m$, and Z_m has precisely one subgroup of order n, Z_n, which is cyclic (Proposition 4.3(**iv**)).

If $\psi : A \to Z_m$ is a homomorphism, and x is a generator for A, then the period of x is an exponent for $\psi(x)$, so that $\psi(x)$, and hence $\psi(A)$, is contained in Z_n. Let y be a generator for Z_n. We have an isomorphism

$$\psi_1 : A \to Z_n$$

such that $\psi_1(x) = y$. For each integer k with $0 \leqq k < n$ we have the homomorphism $k\psi_1$ such that

$$(k\psi_1)(x) = k \cdot \psi_1(x) = \psi_1(kx).$$

In this way we get a cyclic subgroup of $A^\wedge$ consisting of the n elements $k\psi_1$ $(0 \leqq k < n)$. Conversely, any element ψ of $A^\wedge$ is uniquely determined by its effect on the generator x, and must map x on one of the n elements kx $(0 \leqq k < n)$ of Z_n. Hence ψ is equal to one of the maps $k\psi_1$. These maps constitute the full group $A^\wedge$, which is therefore cyclic of order n, generated by ψ_1. This proves our theorem.

In considering the dual group, we take various cyclic groups Z_m. There are many applications where such groups occur, for instance the group of m-th roots of unity in the complex numbers, the subgroup of order m of $\mathbf{Q}/\mathbf{Z}$, etc.

Let A, A' be two abelian groups. A **bilinear** map of $A \times A'$ into an abelian group C is a map

$$A \times A' \to C$$

denoted by

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

having the following property. For each $x \in A$ the function $x' \mapsto \langle x, x' \rangle$ is a homomorphism, and similarly for each $x' \in A'$ the function $x \mapsto \langle x, x' \rangle$ is a homomorphism.

As a special case of a bilinear map, we have the one given by

$$A \times \operatorname{Hom}(A, C) \to C$$

which to each pair (x, f) with $x \in A$ and $f \in \operatorname{Hom}(A, C)$ associates the element $f(x)$ in C.

A bilinear map is also called a **pairing**.

An element $x \in A$ is said to be **orthogonal** (or **perpendicular**) to a subset S' of A' if $\langle x, x' \rangle = 0$ for all $x' \in S'$. It is clear that the set of $x \in A$ orthogonal to S' is a subgroup of A. We make similar definitions for elements of A', orthogonal to subsets of A.

The **kernel** of our bilinear map on the left is the subgroup of A which is orthogonal to all of A'. We define its kernel on the right similarly.

Given a bilinear map $A \times A' \to C$, let B, B' be the respective kernels of our bilinear map on the left and right. An element x' of A' gives rise to an element of $\operatorname{Hom}(A, C)$ given by $x \mapsto \langle x, x' \rangle$, which we shall denote by $\psi_{x'}$. Since $\psi_{x'}$ vanishes on B we see that $\psi_{x'}$ is in fact a homomorphism of A/B into C.

Furthermore, $\psi_{x'} = \psi_{y'}$ if x', y' are elements of A' such that

$$x' \equiv y' \pmod{B'}.$$

Hence ψ is in fact a homomorphism

$$0 \to A'/B' \to \mathrm{Hom}(A/B, C),$$

which is injective since we defined B' to be the group orthogonal to A. Similarly, we get an injective homomorphism

$$0 \to A/B \to \mathrm{Hom}(A'/B', C).$$

Assume that C is cyclic of order m. Then for any $x' \in A'$ we have

$$m\psi_{x'} = \psi_{mx'} = 0,$$

whence A'/B' has exponent m. Similarly, A/B has exponent m.

Theorem 9.2. *Let $A \times A' \to C$ be a bilinear map of two abelian groups into a cyclic group C of order m. Let B, B' be its respective kernels on the left and right. Assume that A'/B' is finite. Then A/B is finite, and A'/B' is isomorphic to the dual group of A/B (under our map ψ).*

Proof. The injection of A/B into $\mathrm{Hom}(A'/B', C)$ shows that A/B is finite. Furthermore, we get the inequalities

$$\mathrm{ord}\, A/B \leqq \mathrm{ord}(A'/B')^\wedge = \mathrm{ord}\, A'/B'$$

and

$$\mathrm{ord}\, A'/B' \leqq \mathrm{ord}(A/B)^\wedge = \mathrm{ord}\, A/B.$$

From this it follows that our map ψ is bijective, hence an isomorphism.

Corollary 9.3. *Let A be a finite abelian group, B a subgroup, $A^\wedge$ the dual group, and $B^\perp$ the set of $\varphi \in A^\wedge$ such that $\varphi(B) = 0$. Then we have a natural isomorphism of $A^\wedge/B^\perp$ with $B^\wedge$.*

Proof. This is a special case of Theorem 9.2.

§10. INVERSE LIMIT AND COMPLETION

Consider a sequence of groups $\{G_n\}$ $(n = 0, 1, 2, \ldots)$, and suppose given for all $n \geqq 1$ homomorphisms

$$f_n \colon G_n \to G_{n-1}.$$

Suppose first that these homomorphisms are surjective. We form infinite sequences

$$x = (x_0, x_1, x_2, \ldots) \quad \text{such that } x_{n-1} = f_n(x_n).$$

By the assumption of surjectivity, given $x_n \in G_n$ we can always lift x_n to G_{n+1} via f_{n+1}, so such infinite sequences exist, projecting to any given x_0. We can define multiplication of such sequences componentwise, and it is then immediately verified that the set of sequences is a group, called the **inverse limit** of the family $\{(G_n, f_n)\}$. We denote the inverse limit by $\varprojlim (G_n, f_n)$, or simply lim G_n if the reference to f_n is clear.

Example. Let A be an additive abelian group. Let p be a prime number. Let $p_A: A \to A$ denote multiplication by p. We say that A is ***p*-divisible** if p_A is surjective. We may then form the inverse limit by taking $A_n = A$ for all n, and $f_n = p_A$ for all n. The inverse limit is denoted by $V_p(A)$. We let $T_p(A)$ be the subset of $V_p(A)$ consisting of those infinite sequences as above such that $x_0 = 0$. Let $A[p^n]$ be the kernel of p_A^n. Then

$$T_p(A) = \varprojlim A[p^{n+1}].$$

The group $T_p(A)$ is called the **Tate group** associated with the p-divisible group A. It arose in fairly sophisticated contexts of algebraic geometry due to Deuring and Weil, in the theory of elliptic curves and abelian varieties developed in the 1940s, which are far afield from this book. Interested readers can consult books on those subjects.

The most common p-divisible groups are obtained as follows. First, let A be the subgroup of $\mathbf{Q}/\mathbf{Z}$ consisting of those rational numbers (mod $\mathbf{Z}$) which can be expressed in the form a/p^k with some positive integer k, and $a \in \mathbf{Z}$. Then A is p-divisible.

Second, let $\boldsymbol{\mu}[p^n]$ be the group of p^n-th roots of unity in the complex numbers. Let $\boldsymbol{\mu}[p^\infty]$ be the union of all $\boldsymbol{\mu}[p^n]$ for all n. Then $\boldsymbol{\mu}[p^\infty]$ is p-divisible, and isomorphic to the group A of the preceding paragraph. Thus

$$T_p(\boldsymbol{\mu}) = \varprojlim \boldsymbol{\mu}[p^n].$$

These groups are quite important in number theory and algebraic geometry. We shall make further comments about them in Chapter III, §10, in a broader context.

Example. Suppose given a group G. Let $\{H_n\}$ be a sequence of normal subgroups such that $H_n \supset H_{n+1}$ for all n. Let

$$f_n: G/H_n \to G/H_{n-1}$$

be the canonical homomorphisms. Then we may form the inverse limit $\varprojlim G/H_n$. Observe that G has a natural homomorphism

$$g: G \to \varprojlim G/H_n,$$

which sends an element x to the sequence $(\ldots, x_n, \ldots)$, where $x_n =$ image of x in G/H_n.

Example. Let $G_n = \mathbf{Z}/p^{n+1}\mathbf{Z}$ for each $n \geqq 0$. Let

$$f_n: \mathbf{Z}/p^{n+1}\mathbf{Z} \to \mathbf{Z}/p^n\mathbf{Z}$$

be the canonical homomorphism. Then f_n is surjective, and the limit is called the group of p-**adic integers**, denoted by $\mathbf{Z}_p$. We return to this in Chapter III, §10, where we shall see that $\mathbf{Z}_p$ is also a ring.

After these examples, we want to consider the more general situation when one deals not with a sequence but with a more general type of family of groups, which may not be commutative. We therefore define inverse limits of groups in general.

Let I be a set of indices. Suppose given a relation of partial ordering in I, namely for some pairs (i, j) we have a relation $i \leqq j$ satisfying the conditions: For all i, j, k in I, we have $i \leqq i$; if $i \leqq j$ and $j \leqq k$ then $i \leqq k$; if $i \leqq j$ and $j \leqq i$ then $i = j$. We say that I is **directed** if given $i, j \in I$, there exists k such that $i \leqq k$ and $j \leqq k$. Assume that I is directed. By an **inversely directed family** of groups, we mean a family $\{G_i\}_{i \in I}$ and for each pair $i \leqq j$ a homomorphism

$$f_i^j \colon G_j \to G_i$$

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

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

Let $G = \prod G_i$ be the product of the family. Let Γ be the subset of G consisting of all elements (x_i) with $x_i \in G_i$ such that for all i and $j \geqq i$ we have

$$f_i^j(x_j) = x_i.$$

Then Γ contains the unit element, and is immediately verified to be a subgroup of G. We call Γ the **inverse limit** of the family, and write

$$\Gamma = \varprojlim G_i.$$

Example. Let G be a group. Let $\mathfrak{F}$ be the family of normal subgroups of finite index. If H, K are normal of finite index, then so is $H \cap K$, so $\mathfrak{F}$ is a directed family. We may then form the inverse limit $\varprojlim G/H$ with $H \in \mathfrak{F}$. There is a variation on this theme. Instead of $\mathfrak{F}$, let p be a prime number, and let $\mathfrak{F}_p$ be the family of normal subgroups of finite index equal to a power of p. Then the inverse limit with respect to subgroups $H \in \mathfrak{F}_p$ can also be taken. (Verify that if H, K are normal of finite p-power index, so is their intersection.)

A group which is an inverse limit of finite groups is called **profinite**.

Example from applications. Such inverse limits arise in Galois theory. Let k be a field and let A be an infinite Galois extension. For example, $k = \mathbf{Q}$ and A is an algebraic closure of $\mathbf{Q}$. Let G be the Galois group; that is, the group of automorphisms of A over k. Then G is the inverse limit of the factor groups G/H, where H ranges over the Galois groups of A over K, with K ranging over all finite extensions of k contained in A. See the Shafarevich conjecture in the chapter on Galois theory, Conjecture 14.2 of Chapter VI.

Similarly, consider a compact Riemann surface X of genus $\geqq 2$. Let $p : X' \to X$ be the universal covering space. Let $\mathbf{C}(X) = F$ and $\mathbf{C}(X') = F'$ be the function fields. Then there is an embedding $\pi_1(X) \hookrightarrow \mathrm{Gal}(F'/F)$. It is shown in complex analysis that $\pi_1(X)$ is a free group with one commutator relation. The full Galois group of F'/F is the inverse limit with respect to the subgroups of finite index, as in the above general situation.

Completion of a group

Suppose now that we are given a group G, and first, for simplicity, suppose given a sequence of normal subgroups $\{H_r\}$ with $H_r \supset H_{r+1}$ for all r, and such that these subgroups have finite index. A sequence $\{x_n\}$ in G will be called a **Cauchy sequence** if given H_r there exists N such that for all $m, n \geqq N$ we have $x_n x_m^{-1} \in H_r$. We say that $\{x_n\}$ is a **null sequence** if given r there exists N such that for all $n \geqq N$ we have $x_n \in H_r$. As an exercise, prove that the Cauchy sequences form a group under termwise product, and that the null sequences form a normal subgroup. The factor group is called the **completion** of G (with respect to the sequence of normal subgroups).

Observe that there is a natural homomorphism of G into its completion; namely, an element $x \in G$ maps to the sequence $(x, x, x, \ldots)$ modulo null sequences. The kernel of this homomorphism is the intersection $\bigcap H_r$, so if this intersection is the unit element of G, then the map of G into its completion is an embedding.

Theorem 10.1. *The completion and the inverse limit* $\varprojlim G/H_r$ *are isomorphic under natural mappings.*

Proof. We give the maps. Let $x = \{x_n\}$ be a Cauchy sequence. Given r, for all n sufficiently large, by the definition of Cauchy sequence, the class of x_n mod H_r is independent of n. Let this class be $x(r)$. Then the sequence $(x(1), x(2), \ldots)$ defines an element of the inverse limit. Conversely, given an element $(\bar{x}_1, \bar{x}_2, \ldots)$ in the inverse limit, with $\bar{x}_n \in G/H_n$, let x_n be a representative in G. Then the sequence $\{x_n\}$ is Cauchy. We leave to the reader to verify that the Cauchy sequence $\{x_n\}$ is well-defined modulo null sequences, and that the maps we have defined are inverse isomorphisms between the completion and the inverse limit.

We used sequences and denumerability to make the analogy with the construction of the real numbers clearer. In general, given the family $\mathfrak{F}$, by a Cauchy family we mean a family $\{x_j\}_{j \in J}$ indexed by an arbitrary directed set J, such that for every $H \in \mathfrak{F}$ there exists $j \in J$ such that for all $k, k' \geqq j$ we have $x_k x_{k'}^{-1} \in H$. In practice, one can work with sequences, because groups that arise naturally are such that the set of subgroups of finite index is denumerable. This occurs when the group G is finitely generated.

More generally, a family $\{H_i\}$ of normal subgroups $\subset \mathfrak{F}$ is called **cofinal** in $\mathfrak{F}$ if given $H \in \mathfrak{F}$ there exists i such that $H_i \subset H$. Suppose that there exists such a family which is denumerable; that is, $i = 1, 2, \ldots$ ranges over the positive integers. Then it is an exercise to show that there is an isomorphism

$$\varprojlim_i G/H_i \approx \varprojlim_{H \in \mathfrak{F}} G/H,$$

or equivalently, that the completion of G with respect to the sequence $\{H_i\}$ is "the same" as the completion with respect to the full family $\mathfrak{F}$. We leave this verification to the reader.

The process of completion is frequent in mathematics. For instance, we shall mention completions of rings in Chapter III, §10; and in Chapter XII we shall deal with completions of fields.

§11. CATEGORIES AND FUNCTORS

Before proceeding further, it will now be convenient to introduce some new terminology. We have met already several kinds of objects: sets, monoids, groups. We shall meet many more, and for each such kind of objects we define special kinds of maps between them (e.g. homomorphisms). Some formal behavior will be common to all of these, namely the existence of identity maps of an object onto itself, and the associativity of maps when such maps occur in succession. We introduce the notion of category to give a general setting for all of these.

A **category** $\mathfrak{A}$ consists of a collection of objects $\mathrm{Ob}(\mathfrak{A})$; and for two objects $A, B \in \mathrm{Ob}(\mathfrak{A})$ a set $\mathrm{Mor}(A, B)$ called the set of **morphisms** of A into B; and for three objects $A, B, C \in \mathrm{Ob}(\mathfrak{A})$ a law of composition (i.e. a map)

$$\mathrm{Mor}(B, C) \times \mathrm{Mor}(A, B) \to \mathrm{Mor}(A, C)$$

satisfying the following axioms:

CAT 1. Two sets $\mathrm{Mor}(A, B)$ and $\mathrm{Mor}(A', B')$ are disjoint unless $A = A'$ and $B = B'$, in which case they are equal.

CAT 2. For each object A of $\mathfrak{A}$ there is a morphism $\mathrm{id}_A \in \mathrm{Mor}(A, A)$ which acts as right and left identity for the elements of $\mathrm{Mor}(A, B)$ and $\mathrm{Mor}(B, A)$ respectively, for all objects $B \in \mathrm{Ob}(\mathfrak{A})$.

CAT 3. The law of composition is associative (when defined), i.e. given $f \in \mathrm{Mor}(A, B)$, $g \in \mathrm{Mor}(B, C)$ and $h \in \mathrm{Mor}(C, D)$ then

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

for all objects A, B, C, D of $\mathfrak{A}$.

Here we write the composition of an element g in $\mathrm{Mor}(B, C)$ and an element f in $\mathrm{Mor}(A, B)$ as $g \circ f$, to suggest composition of mappings. In practice, in this book we shall see that most of our morphisms are actually mappings, or closely related to mappings.

The collection of all morphisms in a category $\mathfrak{A}$ will be denoted by $\mathrm{Ar}(\mathfrak{A})$ ("arrows of $\mathfrak{A}$"). We shall sometimes use the symbols "$f \in \mathrm{Ar}(\mathfrak{A})$" to mean that f is a morphism of $\mathfrak{A}$, i.e. an element of some set $\operatorname{Mor}(A, B)$ for some $A, B \in \operatorname{Ob}(\mathfrak{A})$.

By abuse of language, we sometimes refer to the collection of objects as the category itself, if it is clear what the morphisms are meant to be.

An element $f \in \operatorname{Mor}(A, B)$ is also written $f: A \to B$ or

$$A \xrightarrow{f} B.$$

A morphism f is called an **isomorphism** if there exists a morphism $g: B \to A$ such that $g \circ f$ and $f \circ g$ are the identities in $\operatorname{Mor}(A, A)$ and $\operatorname{Mor}(B, B)$ respectively. If $A = B$, then we also say that the isomorphism is an **automorphism**.

A morphism of an object A into itself is called an **endomorphism**. The set of endomorphisms of A is denoted by $\operatorname{End}(A)$. It follows at once from our axioms that $\operatorname{End}(A)$ is a monoid.

Let A be an object of a category $\mathfrak{A}$. We denote by $\operatorname{Aut}(A)$ the set of automorphisms of A. This set is in fact a group, because all of our definitions are so adjusted so as to see immediately that the group axioms are satisfied (associativity, unit element, and existence of inverse). Thus we now begin to see some feedback between abstract categories and more concrete ones.

Examples. Let $\mathcal{S}$ be the category whose objects are sets, and whose morphisms are maps between sets. We say simply that $\mathcal{S}$ is the category of sets. The three axioms **CAT 1, 2, 3** are trivially satisfied.

Let **Grp** be the category of groups, i.e. the category whose objects are groups and whose morphisms are group-homomorphisms. Here again the three axioms are trivially satisfied. Similarly, we have a category of monoids, denoted by **Mon**.

Later, when we define rings and modules, it will be clear that rings form a category, and so do modules over a ring.

It is important to emphasize here that there are categories for which the set of morphisms is not an abelian group. Some of the most important examples are:

The category $\mathcal{C}^0$, whose objects are open sets in $\mathbf{R}^n$ and whose morphisms are continuous maps.

The category $\mathcal{C}^\infty$ with the same objects, but whose morphisms are the C^∞ maps.

The category **Hol**, whose objects are open sets in $\mathbf{C}^n$, and whose morphisms are holomorphic maps. In each case the axioms of a category are verified, because for instance for **Hol**, the composite of holomorphic maps is holomorphic, and similarly for the other types of maps. Thus a C^0-isomorphism is a continuous map $f: U \to V$ which has a continuous inverse $g: V \to U$. Note that a map may be a C^0-isomorphism but not a C^∞-isomorphism. For instance, $x \mapsto x^3$ is a C^0-automorphism of $\mathbf{R}$, but its inverse is not differentiable.

In mathematics one studies manifolds in any one of the above categories. The determination of the group of automorphisms in each category is one of the basic problems of the area of mathematics concerned with that category. In complex analysis, one determines early the group of holomorphic automorphisms of the unit disc as the group of all maps

$$z \mapsto e^{i\theta}\frac{c - z}{1 - \bar{c}z}$$

with θ real and $c \in \mathbf{C}$, $|c| < 1$.

Next we consider the notion of operation in categories. First, observe that if G is a group, then the G-sets form a category, whose morphisms are the maps $f : S \to S'$ such that $f(xs) = xf(s)$ for $x \in G$ and $s \in S$.

More generally, we can now define the notion of an operation of a group G on an object in any category. Indeed, let $\mathfrak{A}$ be a category and $A \in \mathrm{Ob}(\mathfrak{A})$. By an **operation** of G on A we shall mean a homomorphism of G into the group $\mathrm{Aut}(A)$. In practice, an object A is a set with elements, and an automorphism in $\mathrm{Aut}(A)$ operates on A as a set, i.e. induces a permutation of A. Thus, if we have a homomorphism

$$\rho : G \to \mathrm{Aut}(A),$$

then for each $x \in G$ we have an automorphism $\rho(x)$ of A which is a permutation of A.

An operation of a group G on an object A is also called a **representation** of G on A, and one then says that G is **represented** as a group of automorphisms of A.

Examples. One meets representations in many contexts. In this book, we shall encounter representations of a group on finite-dimensional vector spaces, with the theory pushed to some depth in Chapter XVIII. We shall also deal with representations of a group on modules over a ring. In topology and differential geometry, one represents groups as acting on various topological spaces, for instance spheres. Thus if X is a differential manifold, or a topological manifold, and G is a group, one considers all possible homomorphims of G into $\mathrm{Aut}(X)$, where Aut refers to whatever category is being dealt with. Thus G may be represented in the group of C^0-automorphims, or C^∞-automorphisms, or analytic automorphisms. Such topological theories are not independent of the algebraic theories, because by functoriality, an action of G on the manifold induces an action on various algebraic functors (homology, K-functor, whatever), so that topological or differential problems are to some extent analyzable by the functorial action on the associated groups, vector spaces, or modules.

Let A, B be objects of a category $\mathfrak{A}$. Let $\mathrm{Iso}(A, B)$ be the set of isomorphisms of A with B. Then the group $\mathrm{Aut}(B)$ operates on $\mathrm{Iso}(A, B)$ by composition; namely, if $u \in \mathrm{Iso}(A, B)$ and $v \in \mathrm{Aut}(B)$, then $(v, u) \mapsto v \circ u$ gives the operation. If u_0 is one element of $\mathrm{Iso}(A, B)$, then the orbit of u_0 is all of $\mathrm{Iso}(A, B)$, so $v \mapsto v \circ u_0$ is a bijection $\mathrm{Aut}(B) \to \mathrm{Iso}(A, B)$. The inverse mapping is given by $u \mapsto u \circ u_0^{-1}$. This trivial formalism is very basic, and is applied constantly to each one of the classical categories mentioned above. Of course, we also have a similar bijection on the other side, but the group $\mathrm{Aut}(A)$ operates *on the right* of $\mathrm{Iso}(A, B)$ by composition. Furthermore, if $u: A \to B$ is an isomorphism, then $\mathrm{Aut}(A)$ and $\mathrm{Aut}(B)$ are isomorphic under conjugation, namely

$$w \mapsto uwu^{-1} \quad \text{is an isomorphism} \quad \mathrm{Aut}(A) \to \mathrm{Aut}(B).$$

Two such isomorphisms differ by an inner automorphism. One may visualize this system via the following commutative diagram.

$$\begin{tikzcd}
A \arrow[r, "u"] \arrow[d, "w"'] & B \arrow[d, "uwu^{-1}"] \\
A \arrow[r, "u"'] & B
\end{tikzcd}$$

Let $\rho : G \to \mathrm{Aut}(A)$ and $\rho' : G \to \mathrm{Aut}(A')$ be representations of a group G on two objects A and A' in the same category. A **morphism** of ρ into ρ' is a morphism $h: A \to A'$ such that the following diagram is commutative for all $x \in G$:

$$\begin{tikzcd}
A \arrow[r, "h"] \arrow[d, "\rho(x)"'] & A' \arrow[d, "\rho'(x)"] \\
A \arrow[r, "h"'] & A'
\end{tikzcd}$$

It is then clear that representations of a group G in the objects of a category $\mathfrak{A}$ themselves form a category. An **isomorphism of representations** is then an isomorphism $h : A \to A'$ making the above diagram commutative. An isomorphism of representations is often called an equivalence, but I don't like to tamper with the general system of categorical terminology. Note that if h is an isomorphism of representations, then instead of the above commutative diagram, we let $[h]$ be conjugation by h, and we may use the equivalent diagram

$$\begin{tikzcd}
 & \mathrm{Aut}(A) \arrow[dd, "{[h]}"] \\
G \arrow[ur, "\rho"] \arrow[dr, "\rho'"'] & \\
 & \mathrm{Aut}(A')
\end{tikzcd}$$

Consider next the case where $\mathfrak{A}$ is the category of abelian groups, which we may denote by **Ab**. Let A be an abelian group and G a group. Given an operation of G on the abelian group A, i.e. a homomorphism

$$\rho : G \to \mathrm{Aut}(A),$$

let us denote by $x \cdot a$ the element $\rho_x(a)$. Then we see that for all $x, y \in G$, a, $b \in A$, we have:

$$x \cdot (y \cdot a) = (xy) \cdot a, \qquad x \cdot (a + b) = x \cdot a + x \cdot b,$$
$$e \cdot a = a, \qquad x \cdot 0 = 0.$$

We observe that when a group G operates on itself by conjugation, then not only does G operate on itself as a set but also operates on itself as an object in the category of groups, i.e. the permutations induced by the operation are actually group-automorphisms.

Similarly, we shall introduce later other categories (rings, modules, fields) and we have given a general definition of what it means for a group to operate on an object in any one of these categories.

Let $\mathfrak{A}$ be a category. We may take as objects of a new category $\mathfrak{C}$ the morphisms of $\mathfrak{A}$. If $f: A \to B$ and $f': A' \to B'$ are two morphisms in $\mathfrak{A}$ (and thus objects of $\mathfrak{C}$), then we define a **morphism** $f \to f'$ (in $\mathfrak{C}$) to be a pair of morphisms (φ, ψ) in $\mathfrak{A}$ making the following diagram commutative:

$$\begin{tikzcd}
A \arrow[r, "f"] \arrow[d, "\varphi"'] & B \arrow[d, "\psi"] \\
A' \arrow[r, "f'"'] & B'
\end{tikzcd}$$

In that way, it is clear that $\mathfrak{C}$ is a category. Strictly speaking, as with maps of sets, we should index (φ, ψ) by f and f' (otherwise **CAT 1** is not necessarily satisfied), but such indexing is omitted in practice.

There are many variations on this example. For instance, we could restrict our attention to morphisms in $\mathfrak{A}$ which have a fixed object of departure, or those which have a fixed object of arrival.

Thus let A be an object of $\mathfrak{A}$, and let $\mathfrak{A}_A$ be the category whose objects are morphisms

$$f: X \to A$$

in $\mathfrak{A}$, having A as object of arrival. A morphism in $\mathfrak{A}_A$ from $f: X \to A$ to $g: Y \to A$ is simply a morphism

$$h: X \to Y$$

in $\mathfrak{A}$ such that the diagram is commutative:

$$\begin{tikzcd}
X \arrow[rr, "h"] \arrow[dr, "f"'] & & Y \arrow[dl, "g"] \\
& A &
\end{tikzcd}$$

Universal objects

Let $\mathfrak{C}$ be a category. An object P of $\mathfrak{C}$ is called **universally attracting** if there exists a unique morphism of each object of $\mathfrak{C}$ into P, and is called **universally repelling** if for every object of $\mathfrak{C}$ there exists a unique morphism of P into this object.

When the context makes our meaning clear, we shall call objects P as above **universal**. Since a universal object P admits the identity morphism into itself, it is clear that if P, P' are two universal objects in $\mathfrak{C}$, then there exists a unique isomorphism between them.

Examples. Note that the trivial group consisting only of one element is universal (repelling and attracting) in the category of groups. Similarly, in Chapter II on rings, you will see that the integers $\mathbf{Z}$ are universal in the category of rings (universally repelling).

Next let S be a set. Let $\mathfrak{C}$ be the category whose objects are maps $f : S \to A$ of S into abelian groups, and whose morphisms are the obvious ones: If $f : S \to A$ and $f' : S \to A'$ are two maps into abelian groups, then a morphism of f into f' is a (group) homomorphism $g : A \to A'$ such that the usual diagram is commutative, namely $g \circ f = f'$. Then the free abelian group generated by S is universal in this category. This is a reformulation of the properties we have proved about this group.

Let M be a commutative monoid and let $\gamma \colon M \to K(M)$ be the canonical homomorphism of M into its Grothendieck group. Then γ is universal in the category of homomorphisms of M into abelian groups.

Throughout this book in numerous situations, we define universal objects. Aside from products and coproducts which come immediately after these examples, we have direct and inverse limits; the tensor product in Chapter XVI, §1; the alternating product in Chapter XIX, §1; Clifford algebras in Chapter XIX, §4; *ad lib.*

We now turn to the notion of product in an arbitrary category.

Products and coproducts

Let $\mathfrak{A}$ be a category and let A, B be objects of $\mathfrak{A}$. By a **product** of A, B in $\mathfrak{A}$ one means a triple (P, f, g) consisting of an object P in $\mathfrak{A}$ and two morphisms

$$\begin{tikzcd} & P \arrow[dl, "f"'] \arrow[dr, "g"] & \\ A & & B \end{tikzcd}$$

satisfying the following condition: Given two morphisms

$$\varphi : C \to A \quad \text{and} \quad \psi : C \to B$$

in $\mathfrak{A}$, there exists a unique morphism $h : C \to P$ which makes the following diagram commutative:

$$\begin{tikzcd} & C \arrow[ddl, "\varphi"'] \arrow[ddr, "\psi"] \arrow[d, "h"] & \\ & P \arrow[dl, "f"] \arrow[dr, "g"'] & \\ A & & B \end{tikzcd}$$

In other words, $\varphi = f \circ h$ and $\psi = g \circ h$.

More generally, given a family of objects $\{A_i\}_{i\in I}$ in $\mathfrak{A}$, a **product** for this family consists of $(P, \{f_i\}_{i\in I})$, where P is an object in $\mathfrak{A}$ and $\{f_i\}_{i\in I}$ is a family of morphisms

$$f_i : P \to A_i,$$

satisfying the following condition: Given a family of morphisms

$$g_i : C \to A_i,$$

there exists a unique morphism $h : C \to P$ such that $f_i \circ h = g_i$ for all i.

Example. Let $\mathfrak{A}$ be the category of sets, and let $\{A_i\}_{i\in I}$ be a family of sets. Let $A = \prod\limits_{i\in I} A_i$ be their cartesian product, and let p_i: $A \to A_i$ be the projection on the i-th factor. Then $(A, \{p_i\})$ clearly satisfies the requirements of a product in the category of sets.

As a matter of notation, we shall usually write $A \times B$ for the product of two objects in a category, and $\prod\limits_{i\in I} A_i$ for the product of an arbitrary family in a category, following the same notation as in the category of sets.

Example. *Let $\{G_i\}_{i\in I}$ be a family of groups, and let $G = \prod G_i$ be their direct product. Let p_i: $G \to G_i$ be the projection homomorphism. Then these constitute a product of the family in the category of groups.*

Indeed, if $\{g_i : G' \to G_i\}_{i\in I}$ is a family of homomorphisms, there is a unique homomorphism $g : G' \to \prod G_i$ which makes the required diagram commutative. It is the homomorphism such that $g(x')_i = g_i(x')$ for $x' \in G'$ and each $i \in I$.

Let A, B be objects of a category $\mathfrak{A}$. We note that the product of A, B is universal in the category whose objects consist of pairs of morphisms $f : C \to A$ and g: $C \to B$ in $\mathfrak{A}$, and whose morphisms are described as follows. Let f' : $C' \to A$ and g': $C' \to B$ be another pair. Then a morphism from the first pair to the second is a morphism h: $C \to C'$ in $\mathfrak{A}$, making the following diagram commutative:

$$\begin{tikzcd}
 & C \arrow[ld, "f"'] \arrow[d, "h"] \arrow[rd, "g"] & \\
A & C' \arrow[l, "f'"] \arrow[r, "g'"'] & B
\end{tikzcd}$$

The situation is similar for the product of a family $\{A_i\}_{i\in I}$.

We shall also meet the dual notion: Let $\{A_i\}_{i\in I}$ be a family of objects in a category $\mathfrak{A}$. By their **coproduct** one means a pair $(S, \{f_i\}_{i\in I})$ consisting of an object S and a family of morphisms

$$\{f_i : A_i \to S\},$$

satisfying the following property. Given a family of morphisms $\{g_i : A_i \to C\}$, there exists a unique morphism $h : S \to C$ such that $h \circ f_i = g_i$ for all i.

In the product and coproduct, the morphism h will be said to be the morphism **induced** by the family $\{g_i\}$.

Examples. Let $\mathcal{S}$ be the category of sets. *Then coproducts exist*, i.e. every family of objects has a coproduct. For instance, let S, S' be sets. Let T be a set having the same cardinality as S' and disjoint from S. Let $f_1 : S \to S$ be the identity, and $f_2 : S' \to T$ be a bijection. Let U be the union of S and T. Then (U, f_1, f_2) is a coproduct for S, S', viewing f_1, f_2 as maps into U.

Let $\mathcal{S}_0$ be the category of pointed sets. Its objects consist of pairs (S, x) where S is a set and x is an element of S. A morphism of (S, x) into (S', x') in this category is a map $g : S \to S'$ such that $g(x) = x'$. *Then the coproduct of* (S, x) *and* (S', x') *exists in this category*, and can be constructed as follows. Let T be a set whose cardinality is the same as that of S', and such that $T \cap S = \{x\}$. Let $U = S \cup T$, and let

$$f_1 : (S, x) \to (U, x)$$

be the map which induces the identity on S. Let

$$f_2 : (S', x') \to (U, x)$$

be a map sending x' to x and inducing a bijection of $S' - \{x'\}$ on $T - \{x\}$. Then the triple $((U, x), f_1, f_2)$ is a coproduct for (S, x) and (S', x') in the category of pointed sets.

Similar constructions can be made for the coproduct of arbitrary families of sets or pointed sets. The category of pointed sets is especially important in homotopy theory.

Coproducts are universal objects. Indeed, let $\mathcal{A}$ be a category, and let $\{A_i\}$ be a family of objects in $\mathcal{A}$. We now define $\mathcal{C}$. We let objects of $\mathcal{C}$ be the families of morphisms $\{f_i : A_i \to B\}_{i \in I}$ and given two such families,

$$\{f_i : A_i \to B\} \quad \text{and} \quad \{f'_i : A_i \to B'\},$$

we define a morphism from the first into the second to be a morphism $\varphi : B \to B'$ in $\mathcal{A}$ such that $\varphi \circ f_i = f'_i$ for all i. Then a coproduct of $\{A_i\}$ is simply a universal object in $\mathcal{C}$.

The coproduct of $\{A_i\}$ will be denoted by

$$\coprod_{i \in I} A_i.$$

The coproduct of two objects A, B will also be denoted by $A \amalg B$.

By the general uniqueness statement, we see that it is uniquely determined, up to a unique isomorphism. See the comment, top of p. 58.

Example. Let R be the category of commutative rings. Given two such rings A, B one may form the tensor product, and there are natural ring-homomorphisms $A \to A \otimes B$ and $B \to A \otimes B$ such that

$$a \mapsto a \otimes 1 \text{ and } b \mapsto 1 \otimes b \text{ for } a \in A \text{ and } b \in B.$$

Then the tensor product is a coproduct in the category of commutative rings.

Fiber products and coproducts
Pull-backs and push-outs

Let $\mathcal{C}$ be a category. Let Z be an object of $\mathcal{C}$. Then we have a new category, that of objects over Z, denoted by $\mathcal{C}_Z$. The objects of $\mathcal{C}_Z$ are morphisms:

$$f : X \to Z \text{ in } \mathcal{C}$$

A morphism from f to $g : Y \to Z$ in $\mathcal{C}_Z$ is merely a morphism $h : X \to Y$ in $\mathcal{C}$ which makes the following diagram commutative.

$$\begin{tikzcd}
X \arrow[rr, "h"] \arrow[dr, "f"'] & & Y \arrow[dl, "g"] \\
& Z &
\end{tikzcd}$$

A **product** in $\mathcal{C}_Z$ is called the **fiber product** of f and g in $\mathcal{C}$ and is denoted by $X \times {}_Z Y$, together with its natural morphisms on X, Y over Z, which are sometimes not denoted by anything, but which we denote by p_1, p_2.

$$\begin{tikzcd}
& X \times_Z Y \arrow[dl, "p_1"'] \arrow[dr, "p_2"] & \\
X \arrow[dr, "f"'] & & Y \arrow[dl, "g"] \\
& Z &
\end{tikzcd}$$

Fibered products and coproducts exist in the category of abelian groups

The fibered product of two homomorphisms $f : X \to Z$ and $g : Y \to Z$ is the subgroup of $X \times Y$ consisting of all pairs (x, y) such that

$$f(x) = g(y).$$

The coproduct of two homomorphisms $f : Z \to X$ and $g : Z \to Y$ is the factor group $(X \oplus Y)/W$ where W is the subgroup of $X \oplus Y$ consisting of all elements $(f(z), -g(z))$ with $z \in Z$.
We leave the simple verification to the reader (see Exercises 50–56).

In the fiber product diagram, one also calls p_1 the **pull-back** of g by f, and p_2 the **pull-back** of f by g. The fiber product satisfies the following universal mapping property:

Given any object T in $\mathcal{C}$ and morphisms making the following diagram commutative:

$$\begin{tikzcd}
& T \arrow[dl] \arrow[dr] & \\
X \arrow[dr, "f"'] & & Y \arrow[dl, "g"] \\
& Z &
\end{tikzcd}$$

there exists a unique morphism $T \to X \times_Z Y$ making the following diagram commutative:

$$\begin{tikzcd}
 & X \times_Z Y \arrow[dl, "p_1"'] \arrow[dr, "p_2"] & \\
X & T \arrow[u] \arrow[l] \arrow[r] & Y
\end{tikzcd}$$

Dually, we have the notion of **coproduct** in the category of morphisms $f\colon Z \to X$ with a fixed object Z as the object of departure of the morphisms. This category could be denoted by $\mathcal{C}^Z$. We reverse the arrows in the preceding discussion. Given two objects f and $g\colon Z \to Y$ in this category, we have the notion of their coproduct. It is denoted by $X \amalg_Z Y$, with morphisms q_1, q_2, as in the following commutative diagram:

$$\begin{tikzcd}
 & X \amalg_Z Y & \\
X \arrow[ur, "q_1"] & & Y \arrow[ul, "q_2"'] \\
 & Z \arrow[ul, "f"] \arrow[ur, "g"'] &
\end{tikzcd}$$

satisfying the dual universal property of the fiber product. We call it the **fibered coproduct**. We call q_1 the **push-out** of g by f, and q_2 the **push-out** of f by g.

Example. Let $\mathcal{S}$ be the category of sets. Given two maps f, g as above, their product is the set of all pairs $(x, y) \in X \times Y$ such that $f(x) = g(y)$.

Functors

Let $\mathcal{A}$, $\mathcal{B}$ be categories. A **covariant functor** F of $\mathcal{A}$ into $\mathcal{B}$ is a rule which to each object A in $\mathcal{A}$ associates an object $F(A)$ in $\mathcal{B}$, and to each morphism $f\colon A \to B$ associates a morphism $F(f)\colon F(A) \to F(B)$ such that:

FUN 1. For all A in $\mathcal{A}$ we have $F(\mathrm{id}_A) = \mathrm{id}_{F(A)}$.

FUN 2. If $f\colon A \to B$ and $g\colon B \to C$ are two morphisms of $\mathcal{A}$ then

$$F(g \circ f) = F(g) \circ F(f).$$

Example. If to each group G we associate its set (stripped of the group structure) we obtain a functor from the category of groups into the category of sets, provided that we associate with each group-homomorphism itself, viewed only as a set-theoretic map. Such a functor is called a **stripping** functor or **forgetful functor**.

We observe that a functor transforms isomorphisms into isomorphisms, because $f \circ g = \mathrm{id}$ implies $F(f) \circ F(g) = \mathrm{id}$ also.

We can define the notion of a **contravariant functor** from $\mathcal{A}$ into $\mathcal{B}$ by using essentially the same definition, but reversing all arrows $F(f)$, i.e. to each morphism $f\colon A \to B$ the contravariant functor associates a morphism

$$F(f): F(B) \to F(A)$$

(going in the opposite direction), such that, if

$$f: A \to B \quad \text{and} \quad g: B \to C$$

are morphisms in $\mathfrak{A}$, then

$$F(g \circ f) = F(f) \circ F(g).$$

Sometimes a functor is denoted by writing f_* instead of $F(f)$ in the case of a covariant functor, and by writing f^* in the case of a contravariant functor.

Example. The association $S \mapsto F_{\text{ab}}(S)$ is a covariant functor from the category of sets to the category of abelian groups.

Example. The association which to each group associates its completion with respect to the family of subgroups of finite index is a functor from the category of groups to the category of groups.

Example. Let p be a prime number. Let $\mathcal{C}$ be the category of p-divisible abelian groups. The association $A \mapsto T_p(A)$ is a covariant functor of $\mathcal{C}$ into abelian groups (actually $\mathbf{Z}_p$-modules).

Example. Exercise 49 will show you an example of the group of automorphisms of a forgetful functor.

Example. Let **Man** be the category of compact manifolds. Then the homology is a covariant functor from **Man** into graded abelian groups. The cohomology is a contravariant functor into the category of graded algebras (over the ring of coefficients). The product is the cup product. If the cohomology is taken with coefficients in a field of characteristic 0 (for simplicity), then the cohomology commutes with products. Since cohomology is contravariant, this means that the cohomology of a product is the coproduct of the cohomology of the factors. It turns out that the coproduct is the tensor product, with the graded product, which also gives an example of the use of tensor products. See M. GREENBERG and J. HARPER, *Algebraic Topology* (Benjamin-Addison-Wesley), 1981, Chapter 29.

Example. Let $\mathcal{C}$ be the category of pointed topological spaces (satisfying some mild conditions), i.e. pairs (X, x_0) consisting of a space X and a point x_0. In topology one defines the connected sum of such spaces (X, x_0) and (Y, y_0), glueing X, Y together at the selected point. This connected sum is a coproduct in the category of such pairs, where the morphisms are the continuous maps $f: X \to Y$ such that $f(x_0) = y_0$. Let π_1 denote the fundamental group. Then $(X, x_0) \mapsto \pi_1(X, x_0)$ is a covariant functor from $\mathcal{C}$ into the category of groups, commuting with coproducts. (The existence of coproducts in the category of groups will be proved in §12.)

Example. Suppose we have a morphism $f: X \to Y$ in a category $\mathcal{C}$. By **a section** of f, one means a morphism $g: Y \to X$ such that $g \circ f = \mathrm{id}$. Suppose there exists a covariant functor H from this category to groups such that $H(Y) = \{e\}$ and $H(X) \neq \{e\}$. Then there is no section of f. This is immediate from the formula $H(g \circ f) = \mathrm{id}$, and $H(f) =$ trivial homomorphism. In topology one uses the homology functor to show, for instance, that the unit circle X is not a retract of the closed unit disc with respect to the inclusion mapping f. (Topologists use the word "retract" instead of "section".)

Example. Let $\mathfrak{A}$ be a category and A a fixed object in $\mathfrak{A}$. Then we obtain a covariant functor

$$M_A : \mathfrak{A} \to \mathcal{S}$$

by letting $M_A(X) = \mathrm{Mor}(A, X)$ for any object X of $\mathfrak{A}$. If $\varphi : X \to X'$ is a morphism, we let

$$M_A(\varphi) : \mathrm{Mor}(A, X) \to \mathrm{Mor}(A, X')$$

be the map given by the rule

$$g \mapsto \varphi \circ g$$

for any $g \in \mathrm{Mor}(A, X)$,

$$A \xrightarrow{g} X \xrightarrow{\varphi} X'.$$

The axioms **FUN 1** and **FUN 2** are trivially verified.

Similarly, for each object B of $\mathfrak{A}$, we have a contravariant functor

$$M^B : \mathfrak{A} \to \mathcal{S}$$

such that $M^B(Y) = \mathrm{Mor}(Y, B)$. If $\psi : Y' \to Y$ is a morphism, then

$$M^B(\psi) : \mathrm{Mor}(Y, B) \to \mathrm{Mor}(Y', B)$$

is the map given by the rule

$$f \mapsto f \circ \psi$$

for any $f \in \mathrm{Mor}(Y, B)$,

$$Y' \xrightarrow{\psi} Y \xrightarrow{f} B.$$

The preceding two functors are called the **representation functors**.

Example. Let $\mathfrak{A}$ be the category of abelian groups. Fix an abelian group A. The association $X \mapsto \mathrm{Hom}(A, X)$ is a covariant functor from $\mathfrak{A}$ into itself. The association $X \mapsto \mathrm{Hom}(X, A)$ is a contravariant functor of $\mathfrak{A}$ into itself.

Example. We assume you know about the tensor product. Let A be a commutative ring. Let M be an A-module. The association $X \mapsto M \otimes X$ is a covariant functor from the category of A-modules into itself.

Observe that products and coproducts were defined in a way compatible with the representation functor into the category of sets. Indeed, given a product P

of two objects A and B, then for every object X the set Mor(X, P) is a product of the sets Mor(X, A) and Mor(X, B) in the category of sets. This is merely a reformulation of the defining property of products in arbitrary categories. The system really works.

Let $\mathcal{A}$, $\mathcal{B}$ be two categories. The functors of $\mathcal{A}$ into $\mathcal{B}$ (say covariant, and in one variable) can be viewed as the objects of a category, whose morphisms are defined as follows. Let L, M be two such functors. A **morphism** $H: L \to M$ (also called a **natural transformation**) is a rule which to each object X of $\mathcal{A}$ associates a morphism

$$H_X: L(X) \to M(X)$$

such that for any morphism $f: X \to Y$ the following diagram is commutative:

$$\begin{tikzcd}
L(X) \arrow[r, "H_X"] \arrow[d, "L(f)"'] & M(X) \arrow[d, "M(f)"] \\
L(Y) \arrow[r, "H_Y"'] & M(Y)
\end{tikzcd}$$

We can therefore speak of **isomorphisms** of functors. A functor is **representable** if it is isomorphic to a representation functor as above.

As Grothendieck pointed out, one can use the representation functor to transport the notions of certain structures on sets to arbitrary categories. For instance, let $\mathcal{A}$ be a category and G an object of $\mathcal{A}$. We say that G is a **group object** in $\mathcal{A}$ if for each object X of $\mathcal{A}$ we are given a group structure on the set Mor(X, G) in such a way that the association

$$X \mapsto \mathrm{Mor}(X, G)$$

is functorial (i.e. is a functor from $\mathcal{A}$ into the category of groups). One sometimes denotes the set Mor(X, G) by $G(X)$, and thinks of it as the set of points of G in X. To justify this terminology, the reader is referred to Chapter IX, §2.

Example. Let **Var** be the category of projective non-singular varieties over the complex numbers. To each object X in **Var** one can associate various groups, e.g. Pic(X) (the group of divisor classes for rational equivalence), which is a contravariant functor into the category of abelian groups. Let $\mathrm{Pic}_0(X)$ be the subgroup of classes algebraically equivalent to 0. Then Pic_0 is representable.

In the fifties and sixties Grothendieck was the one who emphasized the importance of the representation functors, and the possibility of transposing to any category notions from more standard categories by means of the representation functors. He himself proved that a number of important functors in algebraic geometry are representable.

§12. FREE GROUPS

We now turn to the coproduct in the category of groups. First a remark. Let $G = \prod G_i$ be a direct product of groups.

We observe that each G_j admits an injective homomorphism into the product, on the j-th component, namely the map $\lambda_j : G_j \to \prod_i G_i$ such that for x in G_j, the i-th component of $\lambda_j(x)$ is the unit element of G_i if $i \neq j$, and is equal to x itself if $i = j$. This embedding will be called the **canonical** one. But we still don't have a coproduct of the family, because the factors commute with each other. To get a coproduct one has to work somewhat harder.

Let G be a group and S a subset of G. We recall that G is **generated** by S if every element of G can be written as a finite product of elements of S and their inverses (the empty product being always taken as the unit element of G). Elements of S are then called **generators**. If there exists a finite set of generators for G we call G **finitely generated**. If S is a set and $\varphi : S \to G$ is a map, we say that φ **generates** G if its image generates G.

Let S be a set, and $f : S \to F$ a map into a group. Let $g : S \to G$ be another map. If $f(S)$ (or as we also say, f) generates F, then it is obvious that there exists at most one homomorphism ψ of F into G which makes the following diagram commutative:

$$\begin{tikzcd}
S \arrow[rr, "f"] \arrow[dr, "g"'] & & F \arrow[dl, "\psi"] \\
& G &
\end{tikzcd}$$

We now consider the category $\mathfrak{C}$ whose objects are the maps of S into groups. If $f : S \to G$ and $f' : S \to G'$ are two objects in this category, we define a morphism from f to f' to be a homomorphism $\varphi : G \to G'$ such that $\varphi \circ f = f'$, i.e. the diagram is commutative:

$$\begin{tikzcd}
& G \arrow[dd, "\varphi"] \\
S \arrow[ur, "f"] \arrow[dr, "f'"'] & \\
& G'
\end{tikzcd}$$

By a **free group** determined by S, we shall mean a universal element in this category.

Proposition 12.1. *Let S be a set. Then there exists a free group (F, f) determined by S. Furthermore, f is injective, and F is generated by the image of f.*

Proof. (I owe this proof to J. Tits.) We begin with a lemma.

Lemma 12.2. *There exists a set I and a family of groups $\{G_i\}_{i \in I}$ such that, if $g\colon S \to G$ is a map of S into a group G, and g generates G, then G is isomorphic to some G_i.*

Proof. This is a simple exercise in cardinalities, which we carry out. If S is finite, then G is finite or denumerable. If S is infinite, then the cardinality of G is $\leqq$ the cardinality of S because G consists of finite products of elements of $g(S)$. Let T be a set which is infinite denumerable if S is finite, and has the same cardinality as S if S is infinite. For each non-empty subset H of T, let Γ_H be the set of group structures on H. For each $\gamma \in \Gamma_H$, let H_γ be the set H, together with the group structure γ. Then the family $\{H_\gamma\}$ for $\gamma \in \Gamma_H$ and H ranging over subsets of T is the desired family.

We return to the proof of the proposition. For each $i \in I$ we let M_i be the set of mappings of S into G_i. For each map $\varphi \in M_i$, we let $G_{i,\varphi}$ be the set-theoretic product of G_i and the set with one element $\{\varphi\}$, so that $G_{i,\varphi}$ is the "same" group as G_i indexed by φ. We let

$$F_0 = \prod_{i \in I} \prod_{\varphi \in M_i} G_{i,\varphi}$$

be the Cartesian product of the groups $G_{i,\varphi}$. We define a map

$$f_0 : S \to F_0$$

by sending S on the factor $G_{i,\varphi}$ by means of φ itself. We contend that given a map $g\colon S \to G$ of S into a group G, there exists a homomorphism $\psi_*\colon F_0 \to G$ making the usual diagram commutative:

$$\begin{tikzcd}
 & F_0 \arrow[dd, "\psi_*"] \\
S \arrow[ur, "f_0"] \arrow[dr, "g"'] & \\
 & G
\end{tikzcd}$$

That is, $\psi_* \circ f_0 = g$. To prove this, we may assume that g generates G, simply by restricting our attention to the subgroup of G generated by the image of g. By the lemma, there exists an isomorphism $\lambda\colon G \to G_i$ for some i, and $\lambda \circ g$ is an element ψ of M_i. We let $\pi_{i,\psi}$ be the projection on the (i, ψ) factor, and we let $\psi_* = \lambda^{-1} \circ \pi_{i,\psi}$. Then the map ψ_* makes the following diagram commutative.

$$\begin{tikzcd}
S \arrow[r, "f_0"] \arrow[d, "g"'] & F_0 \arrow[d, "\pi_{i,\psi}"] \arrow[dl, "\psi_*"'] \\
G \arrow[r, "\lambda"'] & G_{i,\psi}
\end{tikzcd}$$

We let F be the subgroup of F_0 generated by the image of f_0, and we let f simply be equal to f_0, viewed as a map of S into F. We let g_* be the restriction of ψ_* to F. In this way, we see at once that the map g_* is the unique one making our diagram commutative, and thus that (F, f) is the required free group. Furthermore, it is clear that f is injective.

For each set S we select one free group determined by S, and denote it by $(F(S), f_S)$ or briefly by $F(S)$. It is generated by the image of f_S. One may view S as contained in $F(S)$, and the elements of S are called **free** generators of $F(S)$. If $g: S \to G$ is a map, we denote by $g_*: F(S) \to G$ the homomorphism realizing the universality of our free group $F(S)$.

If $\lambda: S \to S'$ is a map of one set into another, we let $F(\lambda): F(S) \to F(S')$ be the map $(f_{S'} \circ \lambda)_*$.

$$\begin{tikzcd}
S \arrow[r, "f_S"] \arrow[d, "\lambda"'] \arrow[dr] & F(S) \arrow[d, "\lambda_* = F(\lambda)"] \\
S' \arrow[r, "f_{S'}"'] & F(S')
\end{tikzcd}$$

Then we may regard F as a functor from the category of sets to the category of groups (the functorial properties are trivially verified, and will be left to the reader).

If λ is surjective, then $F(\lambda)$ is also surjective.

We again leave the proof to the reader.

If two sets S, S' have the same cardinality, then they are isomorphic in the category of sets (an isomorphism being in this case a bijection!), and hence $F(S)$ is isomorphic to $F(S')$. If S has n elements, we call $F(S)$ the **free group on n generators**.

Let G be a group, and let S be the same set as G (i.e. G viewed as a set, without group structure). We have the identity map $g: S \to G$, and hence a surjective homomorphism

$$g_*: F(S) \to G$$

which will be called **canonical**. Thus every group is a factor group of a free group.

One can also construct groups by what is called **generators and relations**. Let S be a set, and $F(S)$ the free group. We assume that $f: S \to F(S)$ is an inclusion. Let R be a set of elements of $F(S)$. Each element of R can be written as a finite product

$$\prod_{\nu=1}^{n} x_\nu$$

where each x_ν is an element of S or an inverse of an element of S. Let N be the smallest normal subgroup of $F(S)$ containing R, i.e. the intersection of all normal subgroups of $F(S)$ containing R. Then $F(S)/N$ will be called the group **determined by the generators S and the relations R**.

Example. One shows easily that the group determined by one generator a, and the relation $\{a^2\}$, has order 2.

The canonical homomorphism $\varphi: F(S) \to F(S)/N$ satisfies the universal mapping property for homomorphisms ψ of $F(S)$ into groups G such that $\psi(x) = e$ for all $x \in R$. In view of this, one sometimes calls the group $F(S)/N$ the group determined by the generators S, and the relations $x = e$ (for all $x \in R$). For instance, the group in the preceding example would be called the group determined by the generator a, and the relation $a^2 = e$.

Let G be a group generated by a finite number of elements, and satisfying the relation $x^2 = e$ for all $x \in G$. What does G look like? It is easy to show that G is commutative. Then one can view G as a vector space over $\mathbf{Z}/2\mathbf{Z}$, so G is determined by its cardinality, up to isomorphism.

In Exercises 34 and 35, you will prove that there exist certain groups satisfying certain relations and with a given order, so that the group presented with these generators and relations can be completely determined. *A priori*, it is not even clear if a group given by generators and relations is finite. Even if it is finite, one does not know its order *a priori*. To show that a group of certain order exists, one has to use various means, a common means being to represent the group as a group of automorphisms of some object, for instance the symmetries of a geometric object. This will be the method suggested for the groups in Exercises 34 and 35, mentioned above.

Example. Let G be a group. For $x, y \in G$ define $[x, y] = xyx^{-1}y^{-1}$ (the commutator) and ${}^{x}y = xyx^{-1}$ (the conjugate). Then one has the cocycle relation

$$[x, yz] = [x, y]\,{}^{y}[x, z].$$

Furthermore, suppose $x, y, z \in G$ and

$$[x, y] = y, \quad [y, z] = z, \quad [z, x] = x.$$

Then $x = y = z = e$. It is an exercise to prove these assertions, but one sees that certain relations imply that a group generated by x, y, z subject to those relations is necessarily trivial.

Next we give a somewhat more sophisticated example. We assume that the reader knows the basic terminology of fields and matrices as in Chapter XIII, but applied only to 2 × 2 matrices. Thus $SL_2(F)$ denotes the group of 2 × 2 matrices with components in a field F and determinant equal to 1.

Example. $SL_2(F)$. Let F be a field. For $b \in F$ and $a \in F$, $a \neq 0$, we let

$$u(b) = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}, \quad s(a) = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}, \quad \text{and } w = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Then it is immediately verified that:

SL 0. $s(a) = wu(a^{-1})wu(a)wu(a^{-1})$.

SL 1. u is an additive homomorphism.

SL 2. s is a multiplicative homomorphism.

SL 3. $w^2 = s(-1)$.

SL 4. $s(a)u(b)s(a^{-1}) = u(ba^2)$.

Now, conversely, suppose that G is an arbitrary group with generators $u(b)$ ($b \in F$) and w, such that if we define $s(a)$ for $a \neq 0$ by **SL 0**, then the relations **SL 1** through **SL 4** are satisfied. Then **SL 3** and **SL 4** show that $s(-1)$ is in the center, and $w^4 = e$. In addition, one verifies that:

SL 5. $ws(a) = s(a^{-1})w$.

Furthermore, one has the theorem:

Let G be the free group with generators $u(b)$, w and relations **SL 1** *through* **SL 4**, *defining $s(a)$ as in* **SL 0**. *Then the natural homomorphism*

$$G \to SL_2(F)$$

is an isomorphism.

Proofs of all the above statements will be found in my $\mathbf{SL_2(R)}$, Springer Verlag, reprint of Addison-Wesley, 1975, Chapter XI, §2. It takes about a page to carry out the proof.

If $F = \mathbf{Q}_p$ is the field of p-adic numbers, then Ihara [Ih 66] proved that every discrete torsion free subgroup of $SL_2(\mathbf{Q}_p)$ is free. Serre put this theorem in the context of a general theory concerning groups acting on trees [Se 80].

[Ih 66] Y. Ihara, On discrete subgroups of the two by two projective linear group over p-adic fields, *J. Math. Soc. Japan* **18** (1966) pp. 219–235

[Se 80] J.-P. Serre, *Trees*, Springer Verlag 1980

Further examples. For further examples of free group constructions, see Exercises 54 and 56. For examples of free groups occurring (possibly conjecturally) in Galois theory, see Chapter VI, §2, Example 9, and the end of Chapter VI, §14.

Proposition 12.3. *Coproducts exist in the category of groups.*

Proof. Let $\{G_i\}_{i \in I}$ be a family of groups. We let $\mathcal{C}$ be the category whose objects are families of group-homomorphisms

$$\{g_i : G_i \to G\}_{i \in I}$$

and whose morphisms are the obvious ones. We must find a universal element in this category. For each index i, we let S_i be the same set as G_i if G_i is infinite, and we let S_i be denumerable if G_i is finite. We let S be a set having the same cardinality as the set-theoretic disjoint union of the sets S_i (i.e. their coproduct in the category of sets). We let Γ be the set of group structures on S, and for each $\gamma \in \Gamma$, we let Φ_γ be the set of all families of homomorphisms

$$\varphi = \{\varphi_i : G_i \to S_\gamma\}.$$

Each pair (S_γ, φ), where $\varphi \in \Phi_\gamma$, is then a group, using φ merely as an index. We let

$$F_0 = \prod_{\gamma \in \Gamma} \prod_{\varphi \in \Phi_\gamma} (S_\gamma, \varphi),$$

and for each i, we define a homomorphism $f_i : G_i \to F_0$ by prescribing the component of f_i on each factor (S_γ, φ) to be the same as that of φ_i.

Let now $g = \{g_i : G_i \to G\}$ be a family of homomorphisms. Replacing G if necessary by the subgroup generated by the images of the g_i, we see that $\text{card}(G) \leqq \text{card}(S)$, because each element of G is a *finite* product of elements in these images. Embedding G as a factor in a product $G \times S_\gamma$ for some γ, we may assume that $\text{card}(G) = \text{card}(S)$. There exists a homomorphism $g_*: F_0 \to G$ such that

$$g_* \circ f_i = g_i$$

for all i. Indeed, we may assume without loss of generality that $G = S_\gamma$ for some γ and that $g = \psi$ for some $\psi \in \Phi_\gamma$. We let g_* be the projection of F_0 on the factor (S_γ, ψ).

Let F be the subgroup of F_0 generated by the union of the images of the maps f_i for all i. The restriction of g_* to F is the unique homomorphism satisfying $f_i \circ g_* = g_i$ for all i, and we have thus constructed our universal object.

Example. Let G_2 be a cyclic group of order 2 and let G_3 be a cyclic group of order 3. What is the coproduct? The answer is neat. It can be shown that $G_2 \amalg G_3$ is the group generated by two elements S, T with relations $S^2 = 1$, $(ST)^3 = 1$. The groups G_2 and G_3 are embedded in $G_2 \amalg G_3$ by sending G_2 on the cyclic group generated by S and sending G_3 on the cyclic group generated by ST. The group can be represented as follows. Let

$$G = SL_2(\mathbf{Z})/\pm 1.$$

As we have seen in an example of §5, the group G operates on the upper half-plane $\mathfrak{H}$. Let S, T be the maps given by

$$S(z) = -1/z \quad \text{and} \quad T(z) = z + 1.$$

Thus S and T are represented by the matrices

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

and satisfy the relations $S^2 = 1$, $(ST)^3 = 1$. Readers will find a proof of several properties of S, T in Serre's *Course in Arithmetic* (Springer Verlag, 1973, Chapter VII, §1), including the fact that S, T generate G. It is an exercise from there to show that G is the coproduct of G_2 and G_3 as asserted.

Observe that these procedures go directly from the universal definition and construction in the proofs of Proposition 12.1 and Proposition 12.3 to the more explicit representation of the free group or the coproduct as the case may be. One relies on the following proposition.

Proposition 12.4. *Let G be a group and $\{G_i\}_{i\in I}$ a family of subgroups. Assume:*

(a) *The family generates G.*

(b) *If*

$$x = x_{i_1} \cdots x_{i_n} \quad \textit{with } x_{i_\nu} \in G_{i_\nu},\ x_{i_\nu} \neq e \textit{ and } i_\nu \neq i_{\nu+1} \textit{ for all } \nu,$$

then $x \neq e$.

Then the natural homomorphism of the coproduct of the family into G sending G_i on itself by the identity mapping is an isomorphism. In other words, simply put, G is the coproduct of the family of subgroups.

Proof. The homomorphism from the coproduct into G is surjective by the assumption that the family generates G. Suppose an element is in the kernel. Then such an element has a representation

$$x_{i_1} \cdots x_{i_n}$$

as in (b), mapping to the identity in G, so all $x_{i_\nu} = e$ and the element itself is equal to e, whence the homomorphism from the coproduct into G is injective, thereby proving the proposition.

Exercises 54 and 56 mentioned above give one illustration of the way Proposition 12.4 can be used. We now show another way, which we carry out for two subgroups. I am indebted to Eilenberg for the neat arrangement of the proof of the next proposition.

Proposition 12.5. *Let A, B be two groups whose set-theoretic intersection is $\{1\}$. There exists a group $A \circ B$ containing A, B as subgroups, such that $A \cap B = \{1\}$, and having the following property. Every element $\neq 1$ of $A \circ B$ has a unique expression as a product*

$$a_1 \cdots a_n \qquad (n \geqq 1, a_i \neq 1 \text{ all } i)$$

with $a_i \in A$ or $a_i \in B$, and such that if $a_i \in A$ then $a_{i+1} \in B$ and if $a_i \in B$ then $a_{i+1} \in A$.

Proof. Let $A \circ B$ be the set of sequences

$$a = (a_1, \dots, a_n) \qquad (n \geqq 0)$$

such that either $n = 0$, and the sequence is empty or $n \geqq 1$, and then elements in the sequence belong to A or B, are $\neq 1$, and two consecutive elements of the sequence do not belong both to A or both to B. If $b = (b_1, \dots, b_m)$, we define the product ab to be the sequence

$$(a_1, \dots, a_n, b_1, \dots, b_m)$$
$$\text{if} \quad a_n \in A, b_1 \in B \quad \text{or} \quad a_n \in B, b_1 \in A,$$

$$(a_1, \dots, a_n b_1, \dots, b_m)$$
$$\text{if} \quad a_n, b_1 \in A \quad \text{or} \quad a_n, b_1 \in B, \quad \text{and} \quad a_n b_1 \neq 1,$$

$$(a_1, \dots, a_{n-1})(b_2, \dots, b_m) \quad \text{by induction,}$$
$$\text{if} \quad a_n, b_1 \in A \quad \text{or} \quad a_n, b_1 \in B \quad \text{and} \quad a_n b_1 = 1.$$

The case when $n = 0$ or $m = 0$ is included in the first case, and the empty sequence is the unit element of $A \circ B$. Clearly,

$$(a_1, \dots, a_n)(a_n^{-1}, \dots, a_1^{-1}) = \text{unit element},$$

so only associativity need be proved. Let $c = (c_1, \dots, c_r)$.

First consider the case $m = 0$, i.e. b is empty. Then clearly $(ab)c = a(bc)$ and similarly if $n = 0$ or $r = 0$. Next consider the case $m = 1$. Let $b = (x)$ with $x \in A$, $x \neq 1$. We then verify in each possible case that $(ab)c = a(bc)$. These cases are as follows:

$$(a_1, \dots, a_n, x, c_1, \dots, c_r) \qquad \text{if} \quad a_n \in B \quad \text{and} \quad c_1 \in B,$$

$$(a_1, \dots, a_n x, c_1, \dots, c_r) \qquad \text{if} \quad a_n \in A, a_n x \neq 1, c_1 \in B,$$

$$(a_1, \dots, a_n, x c_1, \dots, c_r) \qquad \text{if} \quad a_n \in B, c_1 \in A, x c_1 \neq 1,$$

$$(a_1, \dots, a_{n-1})(c_1, \dots, c_r) \qquad \text{if} \quad a_n = x^{-1} \quad \text{and} \quad c_1 \in B,$$

$$\begin{aligned}
&(a_1, \ldots, a_n)(c_2, \ldots, c_r) && \text{if} \quad a_n \in B \quad \text{and} \quad c_1 = x^{-1},\\
&(a_1, \ldots, a_{n-1}, a_n x c_1, c_2, \ldots, c_r) && \text{if} \quad a_n, c_1 \in A, a_n x c_1 \neq 1,\\
&(a_1, \ldots, a_{n-1})(c_2, \ldots, c_r) && \text{if} \quad a_n, c_1 \in A \quad \text{and} \quad a_n x c_1 = 1.
\end{aligned}$$

If $m > 1$, then we proceed by induction. Write $b = b'b''$ with b' and b'' shorter. Then

$$(ab)c = (a(b'b''))c = ((ab')b'')c = (ab')(b''c),$$

$$a(bc) = a((b'b'')c) = a(b'(b''c)) = (ab')(b''c)$$

as was to be shown.

We have obvious injections of A and B into $A \circ B$, and identifying A, B with their images in $A \circ B$ we obtain a proof of our proposition.

We can prove the similar result for several factors. In particular, we get the following corollary for the free group.

Corollary 12.6. *Let $F(S)$ be the free group on a set S, and let $x_1, \ldots, x_n$ be distinct elements of S. Let $\nu_1, \ldots, \nu_r$ be integers $\neq 0$ and let $i_1, \ldots, i_r$ be integers,*

$$1 \leqq i_1, \ldots, i_r \leqq n$$

such that $i_j \neq i_{j+1}$ for $j = 1, \ldots, r - 1$. Then

$$x_{i_1}^{\nu_1} \cdots x_{i_r}^{\nu_r} \neq 1.$$

Proof. Let $G_1, \ldots, G_n$ be the cyclic groups generated by $x_1, \ldots, x_n$. Let $G = G_1 \circ \cdots \circ G_n$. Let

$$F(S) \to G$$

be the homomorphism sending each x_i on x_i, and all other elements of S on the unit element of G. Our assertion follows at once.

Corollary 12.7. *Let S be a set with n elements $x_1, \ldots, x_n$, $n \geqq 1$. Let $G_1, \ldots, G_n$ be the infinite cyclic groups generated by these elements. Then the map*

$$F(S) \to G_1 \circ \cdots \circ G_n$$

sending each x_i on itself is an isomorphism.

Proof. It is obviously surjective and injective.

Corollary 12.8. *Let $G_1, \ldots, G_n$ be groups with $G_i \cap G_j = \{1\}$ if $i \neq j$. The homomorphism*

$$G_1 \amalg \cdots \amalg G_n \to G_1 \circ \cdots \circ G_n$$

of their coproduct into $G_1 \circ \cdots \circ G_n$ induced by the natural inclusion $G_i \to G_1 \circ \cdots \circ G_n$ is an isomorphism.

Proof. Again, it is obviously injective and surjective.

EXERCISES

1. Show that every group of order $\leqq 5$ is abelian.

2. Show that there are two non-isomorphic groups of order 4, namely the cyclic one, and the product of two cyclic groups of order 2.

3. Let G be a group. A **commutator** in G is an element of the form $aba^{-1}b^{-1}$ with a, $b \in G$. Let G^c be the subgroup generated by the commutators. Then G^c is called the **commutator subgroup**. Show that G^c is normal. Show that any homomorphism of G into an abelian group factors through G/G^c.

4. Let H, K be subgroups of a finite group G with $K \subset N_H$. Show that

$$\#(HK) = \frac{\#(H)\#(K)}{\#(H \cap K)}.$$

5. **Goursat's Lemma.** Let G, G' be groups, and let H be a subgroup of $G \times G'$ such that the two projections $p_1 : H \to G$ and $p_2 : H \to G'$ are surjective. Let N be the kernel of p_2 and N' be the kernel of p_1. One can identify N as a normal subgroup of G, and N' as a normal subgroup of G'. Show that the image of H in $G/N \times G'/N'$ is the graph of an isomorphism

$$G/N \approx G'/N'.$$

6. Prove that the group of inner automorphisms of a group G is normal in $\mathrm{Aut}(G)$.

7. Let G be a group such that $\mathrm{Aut}(G)$ is cyclic. Prove that G is abelian.

8. Let G be a group and let H, H' be subgroups. By a **double coset** of H, H' one means a subset of G of the form HxH'.
   (a) Show that G is a disjoint union of double cosets.
   (b) Let $\{c\}$ be a family of representatives for the double cosets. For each $a \in G$ denote by $[a]H'$ the conjugate $aH'a^{-1}$ of H'. For each c we have a decomposition into ordinary cosets

$$H = \bigcup_c x_c(H \cap [c]H'),$$

   where $\{x_c\}$ is a family of elements of H, depending on c. Show that the elements $\{x_c c\}$ form a family of left coset representatives for H' in G; that is,

$$G = \bigcup_{x_c} \bigcup_{x_c} x_c cH',$$

   and the union is disjoint. (Double cosets will not emerge further until Chapter XVIII.)

9. (a) Let G be a group and H a subgroup of finite index. Show that there exists a normal subgroup N of G contained in H and also of finite index. [*Hint*: If $(G : H) = n$, find a homomorphism of G into S_n whose kernel is contained in H.]
   (b) Let G be a group and let H_1, H_2 be subgroups of finite index. Prove that $H_1 \cap H_2$ has finite index.

10. Let G be a group and let H be a subgroup of finite index. Prove that there is only a finite number of right cosets of H, and that the number of right cosets is equal to the number of left cosets.

11. Let G be a group, and A a normal abelian subgroup. Show that G/A operates on A by conjugation, and in this manner get a homomorphism of G/A into $\mathrm{Aut}(A)$.

Semidirect product

12. Let G be a group and let H, N be subgroups with N normal. Let γ_x be conjugation by an element $x \in G$.

    (a) Show that $x \mapsto \gamma_x$ induces a homomorphism $f\colon H \mapsto \mathrm{Aut}(N)$.

    (b) If $H \cap N = \{e\}$, show that the map $H \times N \to HN$ given by $(x, y) \mapsto xy$ is a bijection, and that this map is an isomorphism if and only if f is trivial, i.e. $f(x) = \mathrm{id}_N$ for all $x \in H$.

    We define G to be the **semidirect product** of H and N if $G = NH$ and $H \cap N = \{e\}$.

    (c) Conversely, let N, H be groups, and let $\psi\colon H \to \mathrm{Aut}(N)$ be a given homomorphism. Construct a semidirect product as follows. Let G be the set of pairs (x, h) with $x \in N$ and $h \in H$. Define the composition law

    $$(x_1, h_1)(x_2, h_2) = (x_1\psi(h_1)x_2, h_1h_2).$$

    Show that this is a group law, and yields a semidirect product of N and H, identifying N with the set of elements $(x, 1)$ and H with the set of elements $(1, h)$.

13. (a) Let H, N be normal subgroups of a finite group G. Assume that the orders of H, N are relatively prime. Prove that $xy = yx$ for all $x \in H$ and $y \in N$, and that $H \times N \approx HN$.

    (b) Let $H_1, \dots, H_r$ be normal subgroups of G such that the order of H_i is relatively prime to the order of H_j for $i \neq j$. Prove that

    $$H_1 \times \dots \times H_r \approx H_1 \cdots H_r.$$

    Example. If the Sylow subgroups of a finite group are normal, then G is the direct product of its Sylow subgroups.

14. Let G be a finite group and let N be a normal subgroup such that N and G/N have relatively prime orders.

    (a) Let H be a subgroup of G having the same order as G/N. Prove that $G = HN$.

    (b) Let g be an automorphism of G. Prove that $g(N) = N$.

Some operations

15. Let G be a finite group operating on a finite set S with $\#(S) \geqq 2$. Assume that there is only one orbit. Prove that there exists an element $x \in G$ which has no fixed point, i.e. $xs \neq s$ for all $s \in S$.

16. Let H be a proper subgroup of a finite group G. Show that G is not the union of all the conjugates of H. (But see Exercise 23 of Chapter XIII.)

17. Let X, Y be finite sets and let C be a subset of $X \times Y$. For $x \in X$ let $\varphi(x) =$ number of elements $y \in Y$ such that $(x, y) \in C$. Verify that

    $$\#(C) = \sum_{x \in X} \varphi(x).$$

Remark. A subset C as in the above exercise is often called a **correspondence**, and $\varphi(x)$ is the number of elements in Y which correspond to a given element $x \in X$.

18. Let S, T be finite sets. Show that $\#\mathrm{Map}(S, T) = (\#T)^{\#(S)}$.

19. Let G be a finite group operating on a finite set S.
    (a) For each $s \in S$ show that

$$\sum_{t \in Gs} \frac{1}{\#(Gt)} = 1.$$

    (b) For each $x \in G$ define $f(x)$ = number of elements $s \in S$ such that $xs = s$. Prove that the number of orbits of G in S is equal to

$$\frac{1}{\#(G)} \sum_{x \in G} f(x).$$

Throughout, p is a prime number.

20. Let P be a p-group. Let A be a normal subgroup of order p. Prove that A is contained in the center of P.

21. Let G be a finite group and H a subgroup. Let P_H be a p-Sylow subgroup of H. Prove that there exists a p-Sylow subgroup P of G such that $P_H = P \cap H$.

22. Let H be a normal subgroup of a finite group G and assume that $\#(H) = p$. Prove that H is contained in every p-Sylow subgroup of G.

23. Let P, P' be p-Sylow subgroups of a finite group G.
    (a) If $P' \subset N(P)$ (normalizer of P), then $P' = P$.
    (b) If $N(P') = N(P)$, then $P' = P$.
    (c) We have $N(N(P)) = N(P)$.

Explicit determination of groups

24. Let p be a prime number. Show that a group of order p^2 is abelian, and that there are only two such groups up to isomorphism.

25. Let G be a group of order p^3, where p is prime, and G is not abelian. Let Z be its center. Let C be a cyclic group of order p.
    (a) Show that $Z \approx C$ and $G/Z \approx C \times C$.
    (b) Every subgroup of G of order p^2 contains Z and is normal.
    (c) Suppose $x^p = 1$ for all $x \in G$. Show that G contains a normal subgroup $H \approx C \times C$.

26. (a) Let G be a group of order pq, where p, q are primes and $p < q$. Assume that $q \not\equiv 1 \bmod p$. Prove that G is cyclic.
    (b) Show that every group of order 15 is cyclic.

27. Show that every group of order < 60 is solvable.

28. Let p, q be distinct primes. Prove that a group of order p^2q is solvable, and that one of its Sylow subgroups is normal.

29. Let p, q be odd primes. Prove that a group of order $2pq$ is solvable.

30. (a) Prove that one of the Sylow subgroups of a group of order 40 is normal.
(b) Prove that one of the Sylow subgroups of a group of order 12 is normal.

31. Determine all groups of order $\leqq 10$ up to isomorphism. In particular, show that a non-abelian group of order 6 is isomorphic to S_3.

32. Let S_n be the permutation group on n elements. Determine the p-Sylow subgroups of S_3, S_4, S_5 for $p = 2$ and $p = 3$.

33. Let σ be a permutation of a finite set I having n elements. Define $e(\sigma)$ to be $(-1)^m$ where

$$m = n - \text{number of orbits of } \sigma.$$

If $I_1, \ldots, I_r$ are the orbits of σ, then m is also equal to the sum

$$m = \sum_{\nu=1}^{r} [\text{card}(I_\nu) - 1].$$

If τ is a transposition, show that $e(\sigma\tau) = -e(\sigma)$ be considering the two cases when i, j lie in the same orbit of σ, or lie in different orbits. In the first case, $\sigma\tau$ has one more orbit and in the second case one less orbit than σ. In particular, the sign of a transposition is -1. Prove that $e(\sigma) = \varepsilon(\sigma)$ is the sign of the permutation.

34. (a) Let n be an even positive integer. Show that there exists a group of order $2n$, generated by two elements σ, τ such that $\sigma^n = e = \tau^2$, and $\sigma\tau = \tau\sigma^{n-1}$. (Draw a picture of a regular n-gon, number the vertices, and use the picture as an inspiration to get σ, τ.) This group is called the **dihedral group**.
(b) Let n be an odd positive integer. Let D_{4n} be the group generated by the matrices

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \zeta & 0 \\ 0 & \zeta^{-1} \end{pmatrix}$$

where ζ is a primitive n-th root of unity. Show that D_{4n} has order $4n$, and give the commutation relations between the above generators.

35. Show that there are exactly two non-isomorphic non-abelian groups of order 8. (One of them is given by generators σ, τ with the relations

$$\sigma^4 = 1, \qquad \tau^2 = 1, \qquad \tau\sigma\tau = \sigma^3.$$

The other is the quaternion group.)

36. Let $\sigma = [123 \cdots n]$ in S_n. Show that the conjugacy class of σ has $(n - 1)!$ elements. Show that the centralizer of σ is the cyclic group generated by σ.

37. (a) Let $\sigma = [i_1 \cdots i_m]$ be a cycle. Let $\gamma \in S_n$. Show that $\gamma\sigma\gamma^{-1}$ is the cycle $[\gamma(i_1) \cdots \gamma(i_m)]$.
(b) Suppose that a permutation σ in S_n can be written as a product of r disjoint cycles, and let $d_1, \ldots, d_r$ be the number of elements in each cycle, in increasing order. Let τ be another permutation which can be written as a product of disjoint cycles, whose cardinalities are $d'_1, \ldots, d'_s$ in increasing order. Prove that σ is conjugate to τ in S_n if and only if $r = s$ and $d_i = d'_i$ for all $i = 1, \ldots, r$.

38. (a) Show that S_n is generated by the transpositions $[12]$, $[13], \ldots, [1n]$.
(b) Show that S_n is generated by the transpositions $[12]$, $[23]$, $[34], \ldots, [n - 1, n]$.

(c) Show that S_n is generated by the cycles $[12]$ and $[123 \dots n]$.

(d) Assume that n is prime. Let $\sigma = [123 \dots n]$ and let $\tau = [rs]$ be any transposition. Show that σ, τ generate S_n.

Let G be a finite group operating on a set S. Then G operates in a natural way on the Cartesian product $S^{(n)}$ for each positive integer n. We define the operation on S to be n-**transitive** if given n distinct elements $(s_1, \dots, s_n)$ and n distinct elements $(s_1', \dots, s_n')$ of S, there exists $\sigma \in G$ such that $\sigma s_i = s_i'$ for all $i = 1, \dots, n$.

39. Show that the action of the alternating group A_n on $\{1, \dots, n\}$ is $(n - 2)$-transitive.

40. Let A_n be the alternating group of even permutations of $\{1, \dots, n\}$. For $j = 1, \dots, n$ let H_j be the subgroup of A_n fixing j, so $H_j \approx A_{n-1}$, and $(A_n : H_j) = n$ for $n \geqq 3$. Let $n \geqq 3$ and let H be a subgroup of index n in A_n.

    (a) Show that the action of A_n on cosets of H by left translation gives an isomorphism A_n with the alternating group of permutations of A_n/H.

    (b) Show that there exists an automorphism of A_n mapping H_1 on H, and that such an automorphism is induced by an inner automorphism of S_n if and only if $H = H_i$ for some i.

41. Let H be a simple group of order 60.

    (a) Show that the action of H by conjugation on the set of its Sylow subgroups gives an imbedding $H \hookrightarrow A_6$.

    (b) Using the preceding exercise, show that $H \approx A_5$.

    (c) Show that A_6 has an automorphism which is not induced by an inner automorphism of S_6.

Abelian groups

42. Viewing $\mathbf{Z}$, $\mathbf{Q}$ as additive groups, show that $\mathbf{Q}/\mathbf{Z}$ is a torsion group, which has one and only one subgroup of order n for each integer $n \geqq 1$, and that this subgroup is cyclic.

43. Let H be a subgroup of a finite abelian group G. Show that G has a subgroup that is isomorphic to G/H.

44. Let $f : A \to A'$ be a homomorphism of abelian groups. Let B be a subgroup of A. Denote by A^f and A_f the image and kernel of f in A respectively, and similarly for B^f and B_f. Show that $(A : B) = (A^f : B^f)(A_f : B_f)$, in the sense that if two of these three indices are finite, so is the third, and the stated equality holds.

45. Let G be a finite cyclic group of order n, generated by an element σ. Assume that G operates on an abelian group A, and let $f, g : A \to A$ be the endomorphisms of A given by

$$f(x) = \sigma x - x \quad \text{and} \quad g(x) = x + \sigma x + \cdots + \sigma^{n-1}x.$$

Define the **Herbrand quotient** by the expression $q(A) = (A_f : A^g)/(A_g : A^f)$, provided both indices are finite. Assume now that B is a subgroup of A such that $GB \subset B$.

(a) Define in a natural way an operation of G on A/B.

(b) Prove that

$$q(A) = q(B)q(A/B)$$

in the sense that if two of these quotients are finite, so is the third, and the stated equality holds.

(c) If A is finite, show that $q(A) = 1$.

(This exercise is a special case of the general theory of Euler characteristics discussed in Chapter XX, Theorem 3.1. After reading this, the present exercise becomes trivial. Why?)

Primitive groups

46. Let G operate on a set S. Let $S = \bigcup S_i$ be a partition of S into disjoint subsets. We say that the partition is **stable** under G if G maps each S_i onto S_j for some j, and hence G induces a permutation of the sets of the partition among themselves. There are two partitions of S which are obviously stable: the partition consisting of S itself, and the partition consisting of the subsets with one element. Assume that G operates transitively, and that S has more than one element. Prove that the following two conditions are equivalent:

    PRIM 1. The only partitions of S which are stable are the two partitions mentioned above.

    PRIM 2. If H is the isotropy group of an element of S, then H is a maximal subgroup of G.

    These two conditions define what is known as a **primitive group**, or more accurately, a **primitive operation** of G on S.

    Instead of saying that the operation of a group G is 2-transitive, one also says that it is **doubly transitive**.

47. Let a finite group G operate transitively and faithfully on a set S with at least 2 elements and let H be the isotropy group of some element s of S. (All the other isotropy groups are conjugates of H.) Prove the following:
    (a) G is doubly transitive if and only if H acts transitively on the complement of s in S.
    (b) G is doubly transitive if and only if $G = HTH$, where T is a subgroup of G of order 2 not contained in H.
    (c) If G is doubly transitive, and $(G : H) = n$, then

    $$\#(G) = d(n - 1)n,$$

    where d is the order of the subgroup fixing two elements. Furthermore, H is a maximal subgroup of G, i.e. G is primitive.

48. Let G be a group acting transitively on a set S with at least 2 elements. For each $x \in G$ let $f(x) =$ number of elements of S fixed by x. Prove:

    (a) $\sum_{x \in G} f(x) = \#(G)$.

    (b) G is doubly transitive if and only if

    $$\sum_{x \in G} f(x)^2 = 2\,\#(G).$$

49. **A group as an automorphism group.** Let G be a group and let $\mathbf{Set}(G)$ be the category of G-sets (i.e. sets with a G-operation). Let $F\colon \mathbf{Set}(G) \to \mathbf{Set}$ be the forgetful functor, which to each G-set assigns the set itself. Show that $\operatorname{Aut}(F)$ is naturally isomorphic to G.

Fiber products and coproducts
Pull-backs and push-outs

50\. (a) Show that fiber products exist in the category of abelian groups. In fact, if X, Y are abelian groups with homomorphisms $f: X \to Z$ and $g: Y \to Z$ show that $X \times_Z Y$ is the set of all pairs (x, y) with $x \in X$ and $y \in Y$ such that $f(x) = g(y)$. The maps p_1, p_2 are the projections on the first and second factor respectively.

(b) Show that the pull-back of a surjective homomorphism is surjective.

51\. (a) Show that fiber products exist in the category of sets.

(b) In any category $\mathcal{C}$, consider the category $\mathcal{C}_Z$ of objects over Z. Let $h: T \to Z$ be a fixed object in this category. Let F be the functor such that

$$F(X) = \mathrm{Mor}_Z(T, X),$$

where X is an object over Z, and Mor_Z denotes morphisms over Z. Show that F transforms fiber products over Z into products in the category of sets. (Actually, once you have understood the definitions, this is tautological.)

52\. (a) Show that push-outs (i.e. fiber coproducts) exist in the category of abelian groups. In this case the fiber coproduct of two homomorphisms f, g as above is denoted by $X \oplus_Z Y$. Show that it is the factor group

$$X \oplus_Z Y = (X \oplus Y)/W,$$

where W is the subgroup consisting of all elements $(f(z), -g(z))$ with $z \in Z$.

(b) Show that the push-out of an injective homomorphism is injective.

Remark. After you have read about modules over rings, you should note that the above two exercises apply to modules as well as to abelian groups.

53\. Let H, G, G' be groups, and let

$$f: H \to G, \qquad g: H \to G'$$

be two homomorphisms. Define the notion of coproduct of these two homomorphisms over H, and show that it exists.

54\. (Tits). Let G be a group and let $\{G_i\}_{i\in I}$ be a family of subgroups generating G. Suppose G operates on a set S. For each $i \in I$, suppose given a subset S_i of S, and let s be a point of $S - \bigcup_i S_i$. Assume that for each $g \in G_i - \{e\}$, we have

$$gS_j \subset S_i \text{ for all } j \neq i, \quad \text{and} \quad g(s) \in S_i \text{ for all } i.$$

Prove that G is the coproduct of the family $\{G_i\}_{i\in I}$. (*Hint*: Suppose a product $g_1 \cdots g_m = \mathrm{id}$ on S. Apply this product to s, and use Proposition 12.4.)

55\. Let $M \in GL_2(\mathbf{C})$ (2×2 complex matrices with non-zero determinant). We let

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \text{ and for } z \in \mathbf{C} \text{ we let } M(z) = \frac{az + b}{cz + d}.$$

If $z = -d/c$ ($c \neq 0$) then we put $M(z) = \infty$. Then you can verify (and you should have seen something like this in a course in complex analysis) that $GL_2(\mathbf{C})$ thus operates on $\mathbf{C} \cup \{\infty\}$. Let λ, λ' be the eigenvalues of M viewed as a linear map on $\mathbf{C}^2$. Let W, W' be the corresponding eigenvectors,

$$W = {}^t(w_1, w_2) \text{ and } W' = {}^t(w_1', w_2').$$

By a **fixed point** of M on $\mathbf{C}$ we mean a complex number z such that $M(z) = z$. *Assume that M has two distinct fixed points* $\neq \infty$.

(a) Show that there cannot be more than two fixed points and that these fixed points are $w = w_1/w_2$ and $w' = w_1'/w_2'$. In fact one may take

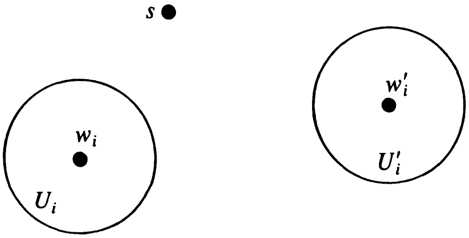
$$W = {}^t(w, 1),\ W' = {}^t(w', 1).$$

(b) Assume that $|\lambda| < |\lambda'|$. Given $z \neq w$, show that

$$\lim_{k \to \infty} M^k(z) = w'.$$

[*Hint*: Let $S = (W, W')$ and consider $S^{-1}M^kS(z) = \alpha^k z$ where $\alpha = \lambda/\lambda'$.]

56. (Tits) Let $M_1, \ldots, M_r \in GL_2(\mathbf{C})$ be a finite number of matrices. Let λ_i, λ_i' be the eigenvalues of M_i. Assume that each M_i has two distinct complex fixed points, and that $|\lambda_i| < |\lambda_i'|$. Also assume that the fixed points for $M_1, \ldots, M_r$ are all distinct from each other. Prove that there exists a positive integer k such that $M_1^k, \ldots, M_r^k$ are the free generators of a free subgroup of $GL_2(\mathbf{C})$. [*Hint*: Let w_i, w_i' be the fixed points of M_i. Let U_i be a small disc centered at w_i and U_i' a small disc centered at w_i'. Let $S_i = U_i \cup U_i'$. Let s be a complex number which does not lie in any S_i. Let $G_i = \langle M_i^k \rangle$. Show that the conditions of Exercise 54 are satisfied for k sufficiently large.].



57. Let G be a group acting on a set X. Let Y be a subset of X. Let G_Y be the subset of G consisting of those elements g such that $gY \cap Y$ is not empty. Let $\overline{G}_Y$ be the subgroup of G generated by G_Y. Then $\overline{G}_Y Y$ and $(G - \overline{G}_Y)Y$ are disjoint. [*Hint*: Suppose that there exist $g_1 \in \overline{G}_Y$ and $g_2 \in G$ but $g_2 \notin \overline{G}_Y$, and elements $y_1, y_2, \in Y$ such that $g_2 y_1 = g_2 y_2$. Then $g_2^{-1} g_1 y_1 = y_2$, so $g_2^{-1} g_1 \in G_Y$ whence $g_2 \in \overline{G}_Y$, contrary to assumption.]

Application. Suppose that $X = GY$, but that X cannot be expressed as a disjoint union as above unless one of the two sets is empty. Then we conclude that $G - \overline{G}_Y$ is empty, and therefore G_Y generates G.

Example 1. Suppose X is a connected topological space, Y is open, and G acts continuously. Then all translates of Y are open, so G is generated by G_Y.

Example 2. Suppose G is a discrete group acting continuously and discretely on X. Again suppose X connected and Y closed, and that any union of translates of Y by elements of G is closed, so again $G - \overline{G}_Y$ is empty, and G_Y generates G.

CHAPTER II

Rings

§1. RINGS AND HOMOMORPHISMS

A **ring** A is a set, together with two laws of composition called multiplication and addition respectively, and written as a product and as a sum respectively, satisfying the following conditions:

RI 1. With respect to addition, A is a commutative group.

RI 2. The multiplication is associative, and has a unit element.

RI 3. For all $x, y, z \in A$ we have

$$(x + y)z = xz + yz \qquad \text{and} \qquad z(x + y) = zx + zy.$$

(This is called **distributivity**.)

As usual, we denote the unit element for addition by 0, and the unit element for multiplication by 1. We do not assume that $1 \neq 0$. We observe that $0x = 0$ for all $x \in A$. *Proof*: We have $0x + x = (0 + 1)x = 1x = x$. Hence $0x = 0$. In particular, if $1 = 0$, then A consists of 0 alone.

For any $x, y \in A$ we have $(-x)y = -(xy)$. *Proof*: We have

$$xy + (-x)y = \big(x + (-x)\big)y = 0y = 0,$$

so $(-x)y$ is the additive inverse of xy.

Other standard laws relating addition and multiplication are easily proved, for instance $(-x)(-y) = xy$. We leave these as exercises.

Let A be a ring, and let U be the set of elements of A which have both a right and left inverse. Then U is a multiplicative group. Indeed, if a has a

S. Lang, *Algebra*

right inverse b, so that $ab = 1$, and a left inverse c, so that $ca = 1$, then $cab = b$, whence $c = b$, and we see that c (or b) is a two-sided inverse, and that c itself has a two-sided inverse, namely a. Therefore U satisfies all the axioms of a multiplicative group, and is called the group of **units** of A. It is sometimes denoted by A^*, and is also called the group of **invertible** elements of A. A ring A such that $1 \neq 0$, and such that every non-zero element is invertible is called a **division ring**.

Note. The elements of a ring which are *left* invertible do not necessarily form a group.

Example. (The Shift Operator). Let E be the set of all sequences

$$a = (a_1, a_2, a_3, \dots)$$

of integers. One can define addition componentwise. Let R be the set of all mappings $f\colon E \to E$ of E into itself such that $f(a + b) = f(a) + f(b)$. The law of composition is defined to be composition of mappings. Then R is a ring. (Proof?) Let

$$T(a_1, a_2, a_3, \dots) = (0, a_1, a_2, a_3, \dots).$$

Verify that T is left invertible but not right invertible.

A ring A is said to be **commutative** if $xy = yx$ for all $x, y \in A$. A commutative division ring is called a **field**. We observe that by definition, a field contains at least two elements, namely 0 and 1.

A subset B of a ring A is called a **subring** if it is an additive subgroup, if it contains the multiplicative unit, and if $x, y \in B$ implies $xy \in B$. If that is the case, then B itself is a ring, the laws of operation in B being the same as the laws of operation in A.

For example, the **center** of a ring A is the subset of A consisting of all elements $a \in A$ such that $ax = xa$ for all $x \in A$. One sees immediately that the center of A is a subring.

Just as we proved general associativity from the associativity for three factors, one can prove general distributivity. If $x, y_1, \dots, y_n$ are elements of a ring A, then by induction one sees that

$$x(y_1 + \cdots + y_n) = xy_1 + \cdots + xy_n.$$

If x_i $(i = 1, \dots, n)$ and y_j $(j = 1, \dots, m)$ are elements of A, then it is also easily proved that

$$\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{j=1}^{m} y_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i y_j.$$

Furthermore, distributivity holds for subtraction, e.g.

$$x(y_1 - y_2) = xy_1 - xy_2.$$

We leave all the proofs to the reader.

Examples. *Let S be a set and A a ring. Let* $\mathrm{Map}(S, A)$ *be the set of mappings of S into A. Then* $\mathrm{Map}(S, A)$ *is a ring if for $f, g \in$* $\mathrm{Map}(S, A)$ *we define*

$$(fg)(x) = f(x)g(x) \qquad \textit{and} \qquad (f + g)(x) = f(x) + g(x)$$

for all $x \in S$. The multiplicative unit is the constant map whose value is the multiplicative unit of A. The additive unit is the constant map whose value is the additive unit of A, namely 0. The verification that $\mathrm{Map}(S, A)$ is a ring under the above laws of composition is trivial and left to the reader.

Let M be an additive abelian group, and let A be the set $\mathrm{End}(M)$ of group-homomorphisms of M into itself. We define addition in A to be the addition of mappings, and we define multiplication to be **composition** of mappings. Then it is trivially verified that A is a ring. Its unit element is of course the identity mapping. In general, A is not commutative.

Readers have no doubt met polynomials over a field previously. These provide a basic example of a ring, and will be defined officially for this book in §3.

Let K be a field. The set of $n \times n$ matrices with components in K is a ring. Its units consist of those matrices which are invertible, or equivalently have a non-zero determinant.

Let S be a set and R the set of real-valued functions on S. Then R is a commutative ring. Its units consist of those functions which are nowhere 0. This is a special case of the ring $\mathrm{Map}(S, A)$ considered above.

The convolution product. We shall now give examples of rings whose product is given by what is called convolution. Let G be a group and let K be a field. Denote by $K[G]$ the set of all formal linear combinations $\alpha = \sum a_x x$ with $x \in G$ and $a_x \in K$, such that all but a finite number of a_x are equal to 0. (See §3, and also Chapter III, §4.) If $\beta = \sum b_x x \in K[G]$, then one can define the product

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

With this product, the **group ring** $K[G]$ is a ring, which will be studied extensively in Chapter XVIII when G is a finite group. Note that $K[G]$ is commutative if and only if G is commutative. The second sum on the right above defines what is called a **convolution product**. If f, g are two functions on a group G, we define their **convolution** $f * g$ by

$$(f * g)(z) = \sum_{xy = z} f(x)g(y).$$

Of course this must make sense. If G is infinite, one may restrict this definition to functions which are 0 except at a finite number of elements. Exercise 12 will give an example (actually on a monoid) when another type of restriction allows for a finite sum on the right.

Example from analysis. In analysis one considers a situation as follows. Let $L^1 = L^1(\mathbf{R})$ be the space of functions which are absolutely integrable.

Given functions f, $g \in L^1$, one defines their **convolution product** $f * g$ by

$$(f * g)(x) = \int_{\mathbf{R}} f(x - y)g(y)\, dy.$$

Then this product satisfies all the axioms of a ring, except that there is no unit element. In the case of the group ring or the convolution of Exercise 12, there is a unit element. (What is it?) Note that the convolution product in the case of $L^1(\mathbf{R})$ is commutative, basically because $\mathbf{R}$ is a commutative additive group. More generally, let G be a locally compact group with a Haar measure μ. Then the convolution product is defined by the similar formula

$$(f * g)(x) = \int_G f(xy^{-1})g(y)\, d\mu(y).$$

After these examples, we return to the general theory of rings.

A **left ideal** $\mathfrak{a}$ in a ring A is a subset of A which is a subgroup of the additive group of A, such that $A\mathfrak{a} \subset \mathfrak{a}$ (and hence $A\mathfrak{a} = \mathfrak{a}$ since A contains 1). To define a right ideal, we require $\mathfrak{a}A = \mathfrak{a}$, and a **two-sided ideal** is a subset which is both a left and a right ideal. A two-sided ideal is called simply an **ideal** in this section. Note that (0) and A itself are ideals.

If A is a ring and $a \in A$, then Aa is a left ideal, called **principal**. We say that a is a generator of $\mathfrak{a}$ (over A). Similarly, AaA is a principal two-sided ideal if we define AaA to be the set of all sums $\sum x_i a y_i$ with x_i, $y_i \in A$. Cf. below the definition of the product of ideals. More generally, let $a_1, \ldots, a_n$ be elements of A. We denote by $(a_1, \ldots, a_n)$ the set of elements of A which can be written in the form

$$x_1 a_1 + \cdots + x_n a_n \quad \text{with} \quad x_i \in A.$$

Then this set of elements is immediately verified to be a left ideal, and $a_1, \ldots, a_n$ are called **generators** of the left ideal.

If $\{\mathfrak{a}_i\}_{i \in I}$ is a family of ideals, then their intersection

$$\bigcap_{i \in I} \mathfrak{a}_i$$

is also an ideal. Similarly for left ideals. Readers will easily verify that if $\mathfrak{a} = (a_1, \ldots, a_n)$, then $\mathfrak{a}$ is the intersection of all left ideals containing the elements $a_1, \ldots, a_n$.

A ring A is said to be **commutative** if $xy = yx$ for all x, $y \in A$. In that case, every left or right ideal is two-sided.

A **commutative** ring such that every ideal is principal and such that $1 \neq 0$ is called a **principal** ring.

Examples. The integers $\mathbf{Z}$ form a ring, which is commutative. Let $\mathfrak{a}$ be an ideal $\neq \mathbf{Z}$ and $\neq 0$. If $n \in \mathfrak{a}$, then $-n \in \mathfrak{a}$. Let d be the smallest integer > 0 lying in $\mathfrak{a}$. If $n \in \mathfrak{a}$ then there exist integers q, r with $0 \leqq r < d$ such that

$$n = dq + r.$$

Since $\mathfrak{a}$ is an ideal, it follows that r lies in $\mathfrak{a}$, hence $r = 0$. Hence $\mathfrak{a}$ consists of all multiples qd of d, with $q \in \mathbf{Z}$, and $\mathbf{Z}$ *is a principal ring.*

A similar example is the ring of polynomials in one variable over a field, as will be proved in Chapter IV, also using the Euclidean algorithm.

Let R be the ring of algebraic integers in a number field K. (For definitions, see Chapter VII.) Then R is not necessarily principal, but let $\mathfrak{p}$ be a prime ideal, and let $R_{\mathfrak{p}}$ be the ring of all elements a/b with $a, b \in R$ and $b \notin \mathfrak{p}$. Then in algebraic number theory, it is shown that $R_{\mathfrak{p}}$ is principal, with one prime ideal $\mathfrak{m}_{\mathfrak{p}}$ consisting of all elements a/b as above but with $a \in \mathfrak{p}$. See Exercises 15, 16, and 17.

An example from analysis. Let A be the set of entire functions on the complex plane. Then A is a commutative ring, and every finitely generated ideal is principal. Given a discrete set of complex numbers $\{z_i\}$ and integers $m_i \geqq 0$, there exists an entire function f having zeros at z_i of multiplicity m_i and no other zeros. Every principal ideal is of the form Af for some such f. The group of units A^* in A consists of the functions which have no zeros. It is a nice exercise in analysis to prove the above statements (using the Weierstrass factorization theorem).

We now return to general notions. Let $\mathfrak{a}$, $\mathfrak{b}$ be ideals of A. We define $\mathfrak{ab}$ to be the set of all sums

$$x_1y_1 + \cdots + x_ny_n$$

with $x_i \in \mathfrak{a}$ and $y_i \in \mathfrak{b}$. Then one verifies immediately that $\mathfrak{ab}$ is an ideal, and that the set of ideals forms a multiplicative monoid, the unit element being the ring itself. This unit element is called the **unit ideal**, and is often written (1). If $\mathfrak{a}$, $\mathfrak{b}$ are left ideals, we define their product $\mathfrak{ab}$ as above. It is also a left ideal, and if $\mathfrak{a}$, $\mathfrak{b}$, $\mathfrak{c}$ are left ideals, then we again have associativity: $(\mathfrak{ab})\mathfrak{c} = \mathfrak{a}(\mathfrak{bc})$.

If $\mathfrak{a}$, $\mathfrak{b}$ are left ideals of A, then $\mathfrak{a} + \mathfrak{b}$ (the sum being taken as additive subgroup of A) is obviously a left ideal. Similarly for right and two-sided ideals. Thus ideals also form a monoid under addition. We also have distributivity: If $\mathfrak{a}_1, \dots, \mathfrak{a}_n$, $\mathfrak{b}$ are ideals of A, then clearly

$$\mathfrak{b}(\mathfrak{a}_1 + \cdots + \mathfrak{a}_n) = \mathfrak{ba}_1 + \cdots + \mathfrak{ba}_n,$$

and similarly on the other side. (However, the set of ideals does not form a ring!)

Let $\mathfrak{a}$ be a left ideal. Define $\mathfrak{a}A$ to be the set of all sums $a_1x_1 + \cdots + a_nx_n$ with $a_i \in \mathfrak{a}$ and $x_i \in A$. Then $\mathfrak{a}A$ is an ideal (two-sided).

Suppose that A is commutative. Let $\mathfrak{a}$, $\mathfrak{b}$ be ideals. Then trivially

$$\mathfrak{ab} \subset \mathfrak{a} \cap \mathfrak{b},$$

but equality does not necessarily hold. However, as an exercise, prove that if $\mathfrak{a} + \mathfrak{b} = A$ then $\mathfrak{ab} = \mathfrak{a} \cap \mathfrak{b}$.

As should be known to the reader, the integers $\mathbf{Z}$ satisfy another property besides every ideal being principal, namely unique factorization into primes.

We shall discuss the general phenomenon in §4. Be it noted here only that if a ring A has the property of unique factorization into prime elements, and p is a prime element, then the ideal (p) is prime, and the ring $R_{(p)}$ (defined as above) is principal. See Exercise 6. Thus principal rings may be obtained in a natural way from rings which are not principal.

As Dedekind found out, some form of unique factorization can be recovered in some cases, replacing unique factorization into prime elements by unique factorization of (non-zero) ideals into prime ideals.

Example. There are cases when the non-zero ideals give rise to a group. Let $\mathfrak{o}$ be a subring of a field K such that every element of K is a quotient of elements of $\mathfrak{o}$; that is, of the form a/b with $a, b \in \mathfrak{o}$ and $b \neq 0$. By a **fractional ideal** $\mathfrak{a}$ we mean a non-zero additive subgroup of K such that $\mathfrak{o}\mathfrak{a} \subset \mathfrak{a}$ (and therefore $\mathfrak{o}\mathfrak{a} = \mathfrak{a}$ since $\mathfrak{o}$ contains the unit element); and such that there exists an element $c \in \mathfrak{o}$, $c \neq 0$, such that $c\mathfrak{a} \subset \mathfrak{o}$. We might say that a fractional ideal has bounded denominator. A **Dedekind ring** is a ring $\mathfrak{o}$ as above such that the fractional ideals form a group under multiplication. As proved in books on algebraic number theory, the ring of algebraic integers in a number field is a Dedekind ring. Do Exercise 14 to get the property of unique factorization into prime ideals. See Exercise 7 of Chapter VII for a sketch of this proof.

If $a \in K$, $a \neq 0$, then $\mathfrak{o}a$ is a fractional ideal, and such ideals are called **principal**. The principal fractional ideals form a subgroup. The factor group is called the **ideal class group**, or **Picard group** of $\mathfrak{o}$, and is denoted by Pic($\mathfrak{o}$). See Exercises 13–19 for some elementary facts about Dedekind rings. It is a basic problem to determine Pic($\mathfrak{o}$) for various Dedekind rings arising in algebraic number theory and function theory. See my book *Algebraic Number Theory* for the beginnings of the theory in number fields. In the case of function theory, one is led to questions in algebraic geometry, notably the study of groups of divisor classes on algebraic varieties and all that this entails. The property that the fractional ideals form a group is essentially associated with the ring having "dimension 1" (which we do not define here). In general one is led into the study of modules under various equivalence relations; see for instance the comments at the end of Chapter III, §4.

We return to the general theory of rings.

By a **ring-homomorphism** one means a mapping $f: A \to B$ where A, B are rings, and such that f is a monoid-homomorphism for the multiplicative structures on A and B, and also a monoid-homomorphism for the additive structure. In other words, f must satisfy:

$$f(a + a') = f(a) + f(a'), \qquad f(aa') = f(a)f(a'),$$
$$f(1) = 1, \qquad f(0) = 0,$$

for all $a, a' \in A$. Its **kernel** is defined to be the kernel of f viewed as additive homomorphism.

The kernel of a ring-homomorphism $f: A \to B$ is an ideal of A, as one verifies at once.

Conversely, let $\mathfrak{a}$ be an ideal of the ring A. We can construct the **factor ring** $A/\mathfrak{a}$ as follows. Viewing A and $\mathfrak{a}$ as additive groups, let $A/\mathfrak{a}$ be the factor group. We define a multiplicative law of composition on $A/\mathfrak{a}$: If $x + \mathfrak{a}$ and $y + \mathfrak{a}$ are two cosets of $\mathfrak{a}$, we define $(x + \mathfrak{a})(y + \mathfrak{a})$ to be the coset $(xy + \mathfrak{a})$. This coset is well defined, for if x_1, y_1 are in the same coset as x, y respectively, then one verifies at once that x_1y_1 is in the same coset as xy. Our multiplicative law of composition is then obviously associative, has a unit element, namely the coset $1 + \mathfrak{a}$, and the distributive law is satisfied since it is satisfied for coset representatives. We have therefore defined a ring structure on $A/\mathfrak{a}$, and the canonical map

$$f: A \to A/\mathfrak{a}$$

is then clearly a ring-homomorphism.

If $g: A \to A'$ is a ring-homomorphism whose kernel contains $\mathfrak{a}$, then there exists a unique ring-homomorphism $g_*: A/\mathfrak{a} \to A'$ making the following diagram commutative:

$$\begin{tikzcd}
A \arrow[rr, "g"] \arrow[dr, "f"'] & & A' \\
& A/\mathfrak{a} \arrow[ur, "g_*"'] &
\end{tikzcd}$$

Indeed, viewing f, g as group-homomorphisms (for the additive structures), there is a unique group-homomorphism g_* making our diagram commutative. We contend that g_* is in fact a ring-homomorphism. We could leave the trivial proof to the reader, but we carry it out in full. If $x \in A$, then $g(x) = g_*f(x)$. Hence for $x, y \in A$,

$$\begin{aligned} g_*(f(x)f(y)) &= g_*(f(xy)) = g(xy) = g(x)g(y) \\ &= g_*f(x)g_*f(y). \end{aligned}$$

Given $\xi, \eta \in A/\mathfrak{a}$, there exist $x, y \in A$ such that $\xi = f(x)$ and $\eta = f(y)$. Since $f(1) = 1$, we get $g_*f(1) = g(1) = 1$, and hence the two conditions that g_* be a multiplicative monoid-homomorphism are satisfied, as was to be shown.

The statement we have just proved is equivalent to saying that the canonical map $f: A \to A/\mathfrak{a}$ is universal in the category of homomorphisms whose kernel contains $\mathfrak{a}$.

Let A be a ring, and denote its unit element by e for the moment. The map

$$\lambda: \mathbf{Z} \to A$$

such that $\lambda(n) = ne$ is a ring-homomorphism (obvious), and its kernel is an ideal (n), generated by an integer $n \geqq 0$. We have a canonical injective homomorphism $\mathbf{Z}/n\mathbf{Z} \to A$, which is a (ring) isomorphism between $\mathbf{Z}/n\mathbf{Z}$ and a subring of A. If $n\mathbf{Z}$ is a prime ideal, then $n = 0$ or $n = p$ for some prime number p. In the first case, A contains as a subring a ring which is isomorphic to $\mathbf{Z}$, and which is often identified with $\mathbf{Z}$. In that case, we say that A has **characteristic** 0. If on the other hand $n = p$, then we say that A has **characteristic** p, and A contains (an isomorphic image of) $\mathbf{Z}/p\mathbf{Z}$ as a subring. We abbreviate $\mathbf{Z}/p\mathbf{Z}$ by $\mathbf{F}_p$.

If K is a field, then K has characteristic 0 or $p > 0$. In the first case, K contains as a subfield an isomorphic image of the rational numbers, and in the second case, it contains an isomorphic image of $\mathbf{F}_p$. In either case, this subfield will be called the **prime field** (contained in K). Since this prime field is the smallest subfield of K containing 1 and has no automorphism except the identity, it is customary to identify it with $\mathbf{Q}$ or $\mathbf{F}_p$ as the case may be. By the **prime ring** (in K) we shall mean either the integers $\mathbf{Z}$ if K has characteristic 0, or $\mathbf{F}_p$ if K has characteristic p.

Let A be a subring of a ring B. Let S be a subset of B commuting with A; in other words we have $as = sa$ for all $a \in A$ and $s \in S$. We denote by $A[S]$ the set of all elements

$$\sum a_{i_1 \cdots i_n} s_1^{i_1} \cdots s_n^{i_n},$$

the sum ranging over a finite number of n-tuples $(i_1, \ldots, i_n)$ of integers $\geqq 0$, and $a_{i_1 \cdots i_n} \in A$, $s_1, \ldots, s_n \in S$. If $B = A[S]$, we say that S is a set of **generators** (or more precisely, **ring generators**) for B over A, or that B is **generated** by S over A. If S is finite, we say that B is **finitely generated as a ring over** A. One might say that $A[S]$ consists of all not-necessarily-commutative polynomials in elements of S with coefficients in A. Note that elements of S may not commute with each other.

Example. The ring of matrices over a field is finitely generated over that field, but matrices don't necessarily commute.

As with groups, we observe that a homomorphism is uniquely determined by its effect on generators. In other words, let $f: A \to A'$ be a ring-homomorphism, and let $B = A[S]$ as above. Then there exists at most one extension of f to a ring-homomorphism of B having prescribed values on S.

Let A be a ring, $\mathfrak{a}$ an ideal, and S a subset of A. We write

$$S \equiv 0 \pmod{\mathfrak{a}}$$

if $S \subset \mathfrak{a}$. If $x, y \in A$, we write

$$x \equiv y \pmod{\mathfrak{a}}$$

if $x - y \in \mathfrak{a}$. If $\mathfrak{a}$ is principal, equal to (a), then we also write

$$x \equiv y \pmod{a}.$$

If $f: A \to A/\mathfrak{a}$ is the canonical homomorphism, then $x \equiv y \pmod{\mathfrak{a}}$ means that $f(x) = f(y)$. The congruence notation is sometimes convenient when we want to avoid writing explicitly the canonical map f.

The factor ring $A/\mathfrak{a}$ is also called a **residue class ring**. Cosets of $\mathfrak{a}$ in A are called **residue classes** modulo $\mathfrak{a}$, and if $x \in A$, then the coset $x + \mathfrak{a}$ is called the **residue class of x modulo $\mathfrak{a}$**.

We have defined the notion of an isomorphism in any category, and so a ring-isomorphism is a ring-homomorphism which has a two-sided inverse. As usual we have the criterion:

A ring-homomorphism $f\colon A \to B$ which is bijective is an isomorphism.

Indeed, there exists a set-theoretic inverse $g\colon B \to A$, and it is trivial to verify that g is a ring-homomorphism.

Instead of saying "ring-homomorphism" we sometimes say simply "homomorphism" if the reference to rings is clear. We note that rings form a category (the morphisms being the homomorphisms).

Let $f\colon A \to B$ be a ring-homomorphism. Then the image $f(A)$ of f is a subring of B. Proof obvious.

It is clear that an injective ring-homomorphism $f\colon A \to B$ establishes a ring-isomorphism between A and its image. Such a homomorphism will be called an **embedding** (of rings).

Let $f\colon A \to A'$ be a ring-homomorphism, and let $\mathfrak{a}'$ be an ideal of A'. Then $f^{-1}(\mathfrak{a}')$ is an ideal $\mathfrak{a}$ in A, and we have an induced injective homomorphism

$$A/\mathfrak{a} \to A'/\mathfrak{a}'.$$

The trivial proof is left to the reader.

Proposition 1.1. *Products exist in the category of rings.*

In fact, let $\{A_i\}_{i \in I}$ be a family of rings, and let $A = \prod A_i$ be their product as additive abelian groups. We define a multiplication in A in the obvious way: If $(x_i)_{i \in I}$ and $(y_i)_{i \in I}$ are two elements of A, we define their product to be $(x_i y_i)_{i \in I}$, i.e. we define multiplication componentwise, just as we did for addition. The multiplicative unit is simply the element of the product whose i-th component is the unit element of A_i. It is then clear that we obtain a ring structure on A, and that the projection on the i-th factor is a ring-homomorphism. Furthermore, A together with these projections clearly satisfies the required universal property.

Note that the usual inclusion of A_i on the i-th factor is *not* a ring-homomorphism because it does not map the unit element e_i of A_i on the unit element of A. Indeed, it maps e_i on the element of A having e_i as i-th component, and 0 $(= 0_i)$ as all other components.

Let A be a ring. Elements x, y of A are said to be **zero divisors** if $x \neq 0$, $y \neq 0$, and $xy = 0$. Most of the rings without zero divisors which we consider will be commutative. In view of this, we define a ring A to be **entire** if $1 \neq 0$, if A is commutative, and if there are no zero divisors in the ring. (Entire rings are also called **integral domains**. However, linguistically, I feel the need for an adjective. "Integral" would do, except that in English, "integral" has been used for "integral over a ring" as in Chapter VII, §1. In French, as in English, two words exist with similar roots: "integral" and "entire". The French have used both words. Why not do the same in English? There is a slight psychological impediment, in that it would have been better if the use of "integral" and "entire" were reversed to fit the long-standing French use. I don't know what to do about this.)

Examples. The ring of integers $\mathbf{Z}$ is without zero divisors, and is therefore entire. If S is a set with at least 2 elements, and A is a ring with $1 \neq 0$, then the ring of mappings $\mathrm{Map}(S, A)$ has zero divisors. (Proof?)

Let m be a positive integer $\neq 1$. The ring $\mathbf{Z}/m\mathbf{Z}$ has zero divisors if and only if m is not a prime number. (Proof left as an exercise.) The ring of $n \times n$ matrices over a field has zero divisors if $n \geqq 2$. (Proof?)

The next criterion is used very frequently.

Let A be an entire ring, and let a, b be non-zero elements of A. Then a, b generate the same ideal if and only if there exists a unit u of A such that $b = au$.

Proof. If such a unit exists we have $Ab = Aua = Aa$. Conversely, assume $Aa = Ab$. Then we can write $a = bc$ and $b = ad$ with some elements $c, d \in A$. Hence $a = adc$, whence $a(1 - dc) = 0$, and therefore $dc = 1$. Hence c is a unit.

§2. COMMUTATIVE RINGS

Throughout this section, we let A denote a commutative ring.

A **prime** ideal in A is an ideal $\mathfrak{p} \neq A$ such that $A/\mathfrak{p}$ is entire. Equivalently, we could say that it is an ideal $\mathfrak{p} \neq A$ such that, whenever $x, y \in A$ and $xy \in \mathfrak{p}$, then $x \in \mathfrak{p}$ or $y \in \mathfrak{p}$. A prime ideal is often called simply a **prime**.

Let $\mathfrak{m}$ be an ideal. We say that $\mathfrak{m}$ is a **maximal** ideal if $\mathfrak{m} \neq A$ and if there is no ideal $\mathfrak{a} \neq A$ containing $\mathfrak{m}$ and $\neq \mathfrak{m}$.

Every maximal ideal is prime.

Proof. Let $\mathfrak{m}$ be maximal and let $x, y \in A$ be such that $xy \in \mathfrak{m}$. Suppose $x \notin \mathfrak{m}$. Then $\mathfrak{m} + Ax$ is an ideal properly containing $\mathfrak{m}$, hence equal to A. Hence we can write

$$1 = u + ax$$

with $u \in \mathfrak{m}$ and $a \in A$. Multiplying by y we find

$$y = yu + axy,$$

whence $y \in \mathfrak{m}$ and $\mathfrak{m}$ is therefore prime.

Let $\mathfrak{a}$ be an ideal $\neq A$. Then $\mathfrak{a}$ is contained in some maximal ideal $\mathfrak{m}$.

Proof. The set of ideals containing $\mathfrak{a}$ and $\neq A$ is inductively ordered by ascending inclusion. Indeed, if $\{\mathfrak{b}_i\}$ is a totally ordered set of such ideals, then $1 \notin \mathfrak{b}_i$ for any i, and hence 1 does not lie in the ideal $\mathfrak{b} = \bigcup \mathfrak{b}_i$, which dominates all $\mathfrak{b}_i$. If $\mathfrak{m}$ is a maximal element in our set, then $\mathfrak{m} \neq A$ and $\mathfrak{m}$ is a maximal ideal, as desired.

The ideal $\{0\}$ is a prime ideal of A if and only if A is entire.

(Proof obvious.)

We defined a **field** K to be a commutative ring such that $1 \neq 0$, and such that the multiplicative monoid of non-zero elements of K is a group (i.e. such that whenever $x \in K$ and $x \neq 0$ then there exists an inverse for x). We note that the only ideals of a field K are K and the zero ideal.

If $\mathfrak{m}$ is a maximal ideal of A, then $A/\mathfrak{m}$ is a field.

Proof. If $x \in A$, we denote by $\bar{x}$ its residue class mod $\mathfrak{m}$. Since $\mathfrak{m} \neq A$ we note that $A/\mathfrak{m}$ has a unit element $\neq 0$. Any non-zero element of $A/\mathfrak{m}$ can be written as $\bar{x}$ for some $x \in A$, $x \notin \mathfrak{m}$. To find its inverse, note that $\mathfrak{m} + Ax$ is an ideal of $A \neq \mathfrak{m}$ and hence equal to A. Hence we can write

$$1 = u + yx$$

with $u \in \mathfrak{m}$ and $y \in A$. This means that $\bar{y}\bar{x} = 1$ (i.e. $= \bar{1}$) and hence that $\bar{x}$ has an inverse, as desired.

Conversely, we leave it as an exercise to the reader to prove that:

If $\mathfrak{m}$ is an ideal of A such that $A/\mathfrak{m}$ is a field, then $\mathfrak{m}$ is maximal.

Let $f\colon A \to A'$ be a homomorphism of commutative rings. Let $\mathfrak{p}'$ be a prime ideal of A', and let $\mathfrak{p} = f^{-1}(\mathfrak{p}')$. Then $\mathfrak{p}$ is prime.

To prove this, let $x, y \in A$, and $xy \in \mathfrak{p}$. Suppose $x \notin \mathfrak{p}$. Then $f(x) \notin \mathfrak{p}'$. But $f(x)f(y) = f(xy) \in \mathfrak{p}'$. Hence $f(y) \in \mathfrak{p}'$, as desired.

As an exercise, prove that if f is surjective, and if $\mathfrak{m}'$ is maximal in A', then $f^{-1}(\mathfrak{m}')$ is maximal in A.

Example. Let $\mathbf{Z}$ be the ring of integers. Since an ideal is also an additive subgroup of $\mathbf{Z}$, every ideal $\neq \{0\}$ is principal, of the form $n\mathbf{Z}$ for some integer $n > 0$ (uniquely determined by the ideal). Let $\mathfrak{p}$ be a prime ideal $\neq \{0\}$, $\mathfrak{p} = n\mathbf{Z}$. Then n must be a prime number, as follows essentially directly from the definition of a prime ideal. Conversely, if p is a prime number, then $p\mathbf{Z}$ is a prime ideal (trivial exercise). Furthermore, $p\mathbf{Z}$ is a maximal ideal. Indeed, suppose $p\mathbf{Z}$ contained in some ideal $n\mathbf{Z}$. Then $p = nm$ for some integer m, whence $n = p$ or $n = 1$, thereby proving $p\mathbf{Z}$ maximal.

If n is an integer, the factor ring $\mathbf{Z}/n\mathbf{Z}$ is called the ring of **integers modulo** n. We also denote

$$\mathbf{Z}/n\mathbf{Z} = \mathbf{Z}(n).$$

If n is a prime number p, then the ring of integers modulo p is in fact a field, denoted by $\mathbf{F}_p$. In particular, the multiplicative group of $\mathbf{F}_p$ is called the group of non-zero integers modulo p. From the elementary properties of groups, we get a standard fact of elementary number theory: If x is an integer $\not\equiv 0 \pmod{p}$, then $x^{p-1} \equiv 1 \pmod{p}$. (For simplicity, it is customary to write mod p instead of mod $p\mathbf{Z}$, and similarly to write mod n instead of mod $n\mathbf{Z}$ for any integer n.) Similarly, given an integer $n > 1$, the units in the ring $\mathbf{Z}/n\mathbf{Z}$ consist of those residue classes mod $n\mathbf{Z}$ which are represented by integers $m \neq 0$ and prime to n. The order of the group of units in $\mathbf{Z}/n\mathbf{Z}$ is called by definition $\varphi(n)$ (where φ is known as the **Euler phi-function**). Consequently, if x is an integer prime to n, then $x^{\varphi(n)} \equiv 1 \pmod{n}$.

Theorem 2.1. (Chinese Remainder Theorem). *Let $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ be ideals of A such that $\mathfrak{a}_i + \mathfrak{a}_j = A$ for all $i \neq j$. Given elements $x_1, \ldots, x_n \in A$, there exists $x \in A$ such that $x \equiv x_i \pmod{\mathfrak{a}_i}$ for all i.*

Proof. If $n = 2$, we have an expression

$$1 = a_1 + a_2$$

for some elements $a_i \in \mathfrak{a}_i$, and we let $x = x_2 a_1 + x_1 a_2$.
For each $i \geqq 2$ we can find elements $a_i \in \mathfrak{a}_1$ and $b_i \in \mathfrak{a}_i$ such that

$$a_i + b_i = 1, \qquad i \geqq 2.$$

The product $\prod_{i=2}^{n} (a_i + b_i)$ is equal to 1, and lies in

$$\mathfrak{a}_1 + \prod_{i=2}^{n} \mathfrak{a}_i,$$

i.e. in $\mathfrak{a}_1 + \mathfrak{a}_2 \cdots \mathfrak{a}_n$. Hence

$$\mathfrak{a}_1 + \prod_{i=2}^{n} \mathfrak{a}_i = A.$$

By the theorem for $n = 2$, we can find an element $y_1 \in A$ such that

$$y_1 \equiv 1 \pmod{\mathfrak{a}_1},$$

$$y_1 = 0 \left(\bmod \prod_{i=2}^{n} \mathfrak{a}_i\right).$$

We find similarly elements $y_2, \ldots, y_n$ such that

$$y_j \equiv 1 \pmod{\mathfrak{a}_j} \qquad \text{and} \qquad y_j \equiv 0 \pmod{\mathfrak{a}_i} \qquad \text{for } i \neq j.$$

Then $x = x_1 y_1 + \cdots + x_n y_n$ satisfies our requirements.

In the same vein as above, we observe that if $\mathfrak{a}_1, \dots, \mathfrak{a}_n$ are ideals of a ring A such that

$$\mathfrak{a}_1 + \cdots + \mathfrak{a}_n = A,$$

and if $\nu_1, \dots, \nu_n$ are positive integers, then

$$\mathfrak{a}_1^{\nu_1} + \cdots + \mathfrak{a}_n^{\nu_n} = A.$$

The proof is trivial, and is left as an exercise.

Corollary 2.2. *Let $\mathfrak{a}_1, \dots, \mathfrak{a}_n$ be ideals of A. Assume that $\mathfrak{a}_i + \mathfrak{a}_j = A$ for $i \neq j$. Let*

$$f\colon A \to \prod_{i=1}^{n} A/\mathfrak{a}_i = (A/\mathfrak{a}_1) \times \cdots \times (A/\mathfrak{a}_n)$$

be the map of A into the product induced by the canonical map of A onto $A/\mathfrak{a}_i$ for each factor. Then the kernel of f is $\bigcap_{i=1}^{n} \mathfrak{a}_i$, and f is surjective, thus giving an isomorphism

$$A/\bigcap \mathfrak{a}_i \xrightarrow{\approx} \prod A/\mathfrak{a}_i.$$

Proof. That the kernel of f is what we said it is, is obvious. The surjectivity follows from the theorem.

The theorem and its corollary are frequently applied to the ring of integers $\mathbf{Z}$ and to distinct prime ideals $(p_1), \dots, (p_n)$. These satisfy the hypothesis of the theorem since they are maximal. Similarly, one could take integers $m_1, \dots, m_n$ which are relatively prime in pairs, and apply the theorem to the principal ideals $(m_1) = m_1\mathbf{Z}, \dots, (m_n) = m_n\mathbf{Z}$. This is the ultraclassical case of the Chinese remainder theorem.

In particular, let m be an integer > 1, and let

$$m = \prod_i p_i^{r_i}$$

be a factorization of m into primes, with exponents $r_i \geqq 1$. Then we have a ring-isomorphism:

$$\mathbf{Z}/m\mathbf{Z} \approx \prod_i \mathbf{Z}/p_i^{r_i}\mathbf{Z}.$$

If A is a ring, we denote as usual by A^* the multiplicative group of invertible elements of A. We leave the following assertions as exercises:

The preceding ring-isomorphism of $\mathbf{Z}/m\mathbf{Z}$ onto the product induces a group-isomorphism

$$(\mathbf{Z}/m\mathbf{Z})^* \approx \prod_i (\mathbf{Z}/p_i^{r_i}\mathbf{Z})^*.$$

In view of our isomorphism, we have

$$\varphi(m) = \prod_i \varphi(p_i^{r_i}).$$

If p is a prime number and r an integer $\geqq 1$, then

$$\varphi(p^r) = (p-1)p^{r-1}.$$

One proves this last formula by induction. If $r = 1$, then $\mathbf{Z}/p\mathbf{Z}$ is a field, and the multiplicative group of that field has order $p - 1$. Let r be $\geqq 1$, and consider the canonical ring-homomorphism

$$\mathbf{Z}/p^{r+1}\mathbf{Z} \to \mathbf{Z}/p^r\mathbf{Z},$$

arising from the inclusion of ideals $(p^{r+1}) \subset (p^r)$. We get an induced group-homomorphism

$$\lambda\colon (\mathbf{Z}/p^{r+1}\mathbf{Z})^* \to (\mathbf{Z}/p^r\mathbf{Z})^*,$$

which is surjective because any integer a which represents an element of $\mathbf{Z}/p^r\mathbf{Z}$ and is prime to p will represent an element of $(\mathbf{Z}/p^{r+1}\mathbf{Z})^*$. Let a be an integer representing an element of $(\mathbf{Z}/p^{r+1}\mathbf{Z})^*$, such that $\lambda(a) = 1$. Then

$$a \equiv 1 \pmod{p^r\mathbf{Z}},$$

and hence we can write

$$a \equiv 1 + xp^r \pmod{p^{r+1}\mathbf{Z}}$$

for some $x \in \mathbf{Z}$. Letting $x = 0, 1, \ldots, p - 1$ gives rise to p distinct elements of $(\mathbf{Z}/p^{r+1}\mathbf{Z})^*$, all of which are in the kernel of λ. Furthermore, the element x above can be selected to be one of these p integers because every integer is congruent to one of these p integers modulo (p). Hence the kernel of λ has order p, and our formula is proved.

Note that the kernel of λ is isomorphic to $\mathbf{Z}/p\mathbf{Z}$. (Proof?)

Application: The ring of endomorphisms of a cyclic group. One of the first examples of a ring is the ring of endomorphisms of an abelian group. In the case of a cyclic group, we have the following complete determination.

Theorem 2.3. *Let A be a cyclic group of order n. For each $k \in \mathbf{Z}$ let $f_k\colon A \to A$ be the endomorphism $x \mapsto kx$ (writing A additively). Then $k \mapsto f_k$ induces a ring isomorphism $\mathbf{Z}/n\mathbf{Z} \approx \mathrm{End}(A)$, and a group isomorphism $(\mathbf{Z}/n\mathbf{Z})^* \approx \mathrm{Aut}(A)$.*

Proof. Recall that the additive group structure on End(A) is simply addition of mappings, and the multiplication is composition of mappings. The fact that $k \mapsto f_k$ is a ring-homomorphism is then a restatement of the formulas

$$1a = a, \qquad (k + k')a = ka + k'a, \qquad \text{and} \qquad (kk')a = k(k'a)$$

for $k, k' \in \mathbf{Z}$ and $a \in A$. If a is a generator of A, then $ka = 0$ if and only if $k \equiv 0 \bmod n$, so $\mathbf{Z}/n\mathbf{Z}$ is embedded in End(A). On the other hand, let $f\colon A \to A$ be an endomorphism. Again for a generator a, we have $f(a) = ka$

for some k, whence $f = f_k$ since every $x \in A$ is of the form ma for some $m \in Z$, and

$$f(x) = f(ma) = mf(a) = mka = kma = kx.$$

This proves the isomorphism $\mathbf{Z}/n\mathbf{Z} \approx \operatorname{End}(A)$. Furthermore, if $k \in (\mathbf{Z}/n\mathbf{Z})^*$ then there exists k' such that $kk' \equiv 1 \bmod n$, so f_k has the inverse $f_{k'}$ and f_k is an automorphism. Conversely, given any automorphism f with inverse g, we know from the first part of the proof that $f = f_k$, $g = g_{k'}$ for some k, k', and $f \circ g = \mathrm{id}$ means that $kk' \equiv 1 \bmod n$, so $k, k' \in (\mathbf{Z}/n\mathbf{Z})^*$. This proves the isomorphism $(\mathbf{Z}/n\mathbf{Z})^* \approx \operatorname{Aut}(A)$.

Note that if A is written as a multiplicative group C, then the map f_k is given by $x \mapsto x^k$. For instance, let $\boldsymbol{\mu}_n$ be the group of n-th roots of unity in $\mathbf{C}$. Then all automorphisms of $\boldsymbol{\mu}_n$ are given by

$$\zeta \mapsto \zeta^k \quad \text{with} \quad k \in (\mathbf{Z}/n\mathbf{Z})^*.$$

§3. POLYNOMIALS AND GROUP RINGS

Although all readers will have met polynomial functions, this section lays the ground work for polynomials in general. One needs polynomials over arbitrary rings in many contexts. For one thing, there are polynomials over a finite field which cannot be identified with polynomial functions in that field. One needs polynomials with integer coefficients, and one needs to reduce these polynomials mod p for primes p. One needs polynomials over arbitrary commutative rings, both in algebraic geometry and in analysis, for instance the ring of polynomial differential operators. We also have seen the example of a ring $B = A[S]$ generated by a set of elements over a ring A. We now give a systematic account of the basic definitions of polynomials over a commutative ring A.

We want to give a meaning to an expression such as

$$a_0 + a_1X + \cdots + a_nX^n,$$

where $a_i \in A$ and X is a "variable". There are several devices for doing so, and we pick one of them. (I picked another in my *Undergraduate Algebra.*) Consider an infinite cyclic group generated by an element X. We let S be the subset consisting of powers X^r with $r \geqq 0$. Then S is a monoid. We define the set of **polynomials** $A[X]$ to be the set of functions $S \to A$ which are equal to 0 except for a finite number of elements of S. For each element $a \in A$ we denote by aX^n the function which has the value a on X^n and the value 0 for all other elements of S. Then it is immediate that a polynomial can be written uniquely as a finite sum

$$a_0X^0 + \cdots + a_nX^n$$

for some integer $n \in \mathbf{N}$ and $a_i \in A$. Such a polynomial is denoted by $f(X)$. The elements $a_i \in A$ are called the **coefficients** of f. We define the product according to the convolution rule. Thus, given polynomials

$$f(X) = \sum_{i=0}^{n} a_iX^i \qquad \text{and} \qquad g(X) = \sum_{j=0}^{m} b_jX^j$$

we define the product to be

$$f(X)g(X) = \sum_{k=0}^{m+n} \left(\sum_{i+j=k} a_ib_j \right) X^k.$$

It is immediately verified that this product is associative and distributive. We shall give the details of associativity in the more general context of a monoid ring below. Observe that there is a unit element, namely $1X^0$. There is also an embedding

$$A \to A[X] \qquad \text{given by} \qquad a \mapsto aX^0.$$

One usually does not distinguish a from its image in $A[X]$, and one writes a instead of aX^0. Note that for $c \in A$ we have then $cf(x) = \sum ca_iX^i$.

Observe that by our definition, we have an equality of polynomials

$$\sum a_iX^i = \sum b_iX^i$$

if and only if $a_i = b_i$ for all i.

Let A be a subring of a commutative ring B. Let $x \in B$. If $f \in A[X]$ is a polynomial, we may then define the associated **polynomial function**

$$f_B \colon B \to B$$

by letting

$$f_B(x) = f(x) = a_0 + a_1x + \cdots + a_nx^n.$$

Given an element $b \in B$, directly from the definition of multiplication of polynomials, we find:

The association

$$\mathrm{ev}_b \colon f \mapsto f(b)$$

is a ring homomorphism of $A[X]$ into B.

This homomorphism is called the **evaluation homomorphism**, and is also said to be obtained by **substituting** b for X in the polynomial. (Cf. Proposition 3.1 below.)

Let $x \in B$. We now see that the subring $A[x]$ of B generated by x over A is the ring of all polynomial values $f(x)$, for $f \in A[X]$. If the evaluation map $f \mapsto f(x)$ gives an isomorphism of $A[X]$ with $A[x]$, then we say that x is

transcendental over A, or that x is a **variable** over A. In particular, X is a variable over A.

Example. Let $\alpha = \sqrt{2}$. Then the set of all real numbers of the form $a + b\alpha$, with $a, b \in \mathbf{Z}$, is a subring of the real numbers, generated by $\sqrt{2}$. Note that α is not transcendental over $\mathbf{Z}$, because the polynomial $X^2 - 2$ lies in the kernel of the evaluation map $f \mapsto f(\sqrt{2})$. On the other hand, it can be shown that $e = 2.718\ldots$ and π are transcendental over $\mathbf{Q}$. See Appendix 1.

Example. Let p be a prime number and let $K = \mathbf{Z}/p\mathbf{Z}$. Then K is a field. Let $f(X) = X^p - X \in K[X]$. Then f is not the zero polynomial. But f_K is the zero function. Indeed, $f_K(0) = 0$. If $x \in K$, $x \neq 0$, then since the multiplicative group of K has order $p - 1$, it follows that $x^{p-1} = 1$, whence $x^p = x$, so $f(x) = 0$. Thus a non-zero polynomial gives rise to the zero function on K.

There is another homomorphism of the polynomial ring having to do with the coefficients. Let

$$\varphi: A \to B$$

be a homomorphism of commutative rings. Then there is an associated homomorphism of the polynomial rings $A[X] \to B[X]$, such that

$$f(X) = \sum a_i X^i \mapsto \sum \varphi(a_i) X^i = (\varphi f)(X).$$

The verification that this mapping is a homomorphism is immediate, and further details will be given below in Proposition 3.2, in a more general context. We call $f \mapsto \varphi f$ the **reduction map**.

Examples. In some applications the map φ may be an isomorphism. For instance, if $f(X)$ has complex coefficients, then its complex conjugate $\bar{f}(X) = \sum \bar{a}_i X^i$ is obtained by applying complex conjugation to its coefficients.

Let $\mathfrak{p}$ be a prime ideal of A. Let $\varphi: A \to A'$ be the canonical homomorphism of A onto $A/\mathfrak{p}$. If $f(X)$ is a polynomial in $A[X]$, then φf will sometimes be called the **reduction of f modulo $\mathfrak{p}$**.

For example, taking $A = \mathbf{Z}$ and $\mathfrak{p} = (p)$ where p is a prime number, we can speak of the polynomial $3X^4 - X + 2$ as a polynomial mod 5, viewing the coefficients $3, -1, 2$ as integers mod 5, i.e. elements of $\mathbf{Z}/5\mathbf{Z}$.

We may now combine the evaluation map and the reduction map to generalize the evaluation map.

Let $\varphi: A \to B$ be a homomorphism of commutative rings. Let $x \in B$. There is a unique homomorphism extending φ

$$A[X] \to B \qquad \textit{such that} \qquad X \mapsto x,$$

and for this homomorphism, $\sum a_i X^i \mapsto \sum \varphi(a_i) x^i$.

The homomorphism of the above statement may be viewed as the composite

$$A[X] \longrightarrow B[X] \xrightarrow{\mathrm{ev}_x} B$$

where the first map applies φ to the coefficients of a polynomial, and the second map is the evaluation at x previously discussed.

Example. In Chapter IX, §2 and §3, we shall discuss such a situation in several variables, when $(\varphi f)(x) = 0$, in which case x is called a **zero** of the polynomial f.

When writing a polynomial $f(X) = \sum_{i=0}^{n} a_i X^i$, if $a_n \neq 0$ then we define n to be the **degree** of f. Thus the degree of f is the smallest integer n such that $a_r = 0$ for $r > n$. If $f = 0$ (i.e. f is the zero polynomial), then by convention we define the degree of f to be $-\infty$. We agree to the convention that

$$-\infty + -\infty = -\infty, \qquad -\infty + n = -\infty, \qquad -\infty < n,$$

for all $n \in \mathbf{Z}$, and no other operation with $-\infty$ is defined. A polynomial of degree 1 is also called a **linear** polynomial. If $f \neq 0$ and $\deg f = n$, then we call a_n the **leading coefficient** of f. We call a_0 its **constant term.**

Let

$$g(X) = b_0 + \cdots + b_m X^m$$

be a polynomial in $A[X]$, of degree m, and assume $g \neq 0$. Then

$$f(X)g(X) = a_0 b_0 + \cdots + a_n b_m X^{m+n}.$$

Therefore:

If we assume that at least one of the leading coefficients a_n or b_m is not a divisor of 0 *in A, then*

$$\deg(fg) = \deg f + \deg g$$

and the leading coefficient of fg is $a_n b_m$. This holds in particular when a_n or b_m is a unit in A, or when A is entire. Consequently, when A is entire, $A[X]$ is also entire.

If f or $g = 0$, then we still have

$$\deg(fg) = \deg f + \deg g$$

if we agree that $-\infty + m = -\infty$ for any integer m.

One verifies trivially that for any polynomials f, $g \in A[X]$ we have

$$\deg(f + g) \leqq \max(\deg f, \deg g),$$

again agreeing that $-\infty < m$ for every integer m.

Polynomials in several variables

We now go to polynomials in several variables. Let A be a subring of a commutative ring B. Let $x_1, \dots, x_n \in B$. For each n-tuple of integers $(\nu_1, \dots, \nu_n) = (\nu) \in \mathbf{N}^n$, we use vector notation, letting $(x) = (x_1, \dots, x_n)$, and

$$M_{(\nu)}(x) = x_1^{\nu_1} \cdots x_n^{\nu_n}.$$

The set of such elements forms a monoid under multiplication. Let $A[x] = A[x_1, \dots, x_n]$ be the subring of B generated by $x_1, \dots, x_n$ over A. Then every element of $A[x]$ can be written as a finite sum

$$\sum a_{(\nu)} M_{(\nu)}(x) \qquad \text{with} \quad a_{(\nu)} \in A.$$

Using the construction of polynomials in one variable repeatedly, we may form the ring

$$A[X_1, \dots, X_n] = A[X_1][X_2] \cdots [X_n],$$

selecting X_n to be a variable over $A[X_1, \dots, X_{n-1}]$. Then every element f of $A[X_1, \dots, X_n] = A[X]$ has a *unique* expression as a finite sum

$$f = \sum_{j=0}^{d_n} f_j(X_1, \dots, X_{n-1}) X_n^j \qquad \text{with} \quad f_j \in A[X_1, \dots, X_{n-1}].$$

Therefore by induction we can write f uniquely as a sum

$$\begin{aligned} f &= \sum_{\nu_n=0}^{d_n} \left(\sum_{\nu_1, \dots, \nu_{n-1}} a_{\nu_1 \cdots \nu_n} X_1^{\nu_1} \cdots X_{n-1}^{\nu_{n-1}} \right) X_n^{\nu_n} \\ &= \sum a_{(\nu)} M_{(\nu)}(X) = \sum a_{(\nu)} X_1^{\nu_1} \cdots X_n^{\nu_n} \end{aligned}$$

with elements $a_{(\nu)} \in A$, which are called the **coefficients** of f. The products

$$M_{(\nu)}(X) = X_1^{\nu_1} \cdots X_n^{\nu_n}$$

will be called **primitive monomials**. Elements of $A[X]$ are called **polynomials** (in n variables). We call $a_{(\nu)}$ its **coefficients.**

Just as in the one-variable case, we have an evaluation map. Given $(x) = (x_1, \dots, x_n)$ and f as above, we define

$$f(x) = \sum a_{(\nu)} M_{(\nu)}(x) = \sum a_{(\nu)} x_1^{\nu_1} \cdots x_n^{\nu_n}.$$

Then the **evaluation map**

$$\mathrm{ev}_{(x)} \colon A[X] \to B \qquad \text{such that} \qquad f \mapsto f(x)$$

is a ring-homomorphism. It may be viewed as the composite of the successive evaluation maps in one variable $X_i \mapsto x_i$ for $i = n, \dots, 1$, because $A[X] \subset B[X]$.

Just as for one variable, if $f(X) \in A[X]$ is a polynomial in n variables, then we obtain a function

$$f_B: B^n \to B \quad \text{by} \quad (x) \mapsto f(x).$$

We say that $f(x)$ is obtained by **substituting** (x) for (X) in f, or by **specializing** (X) to (x). As for one variable, if K is a finite field, and $f \in K[X]$ one may have $f \neq 0$ but $f_K = 0$. Cf. Chapter IV, Theorem 1.4 and its corollaries.

Next let $\varphi: A \to B$ be a homomorphism of commutative rings. Then we have the **reduction map** (generalized in Proposition 3.2 below)

$$f(X) = \sum a_{(\nu)} M_{(\nu)}(X) \mapsto \sum \varphi(a_{(\nu)}) M_{(\nu)}(X) = (\varphi f)(X).$$

We can also compose the evaluation and reduction. An element $(x) \in B^n$ is called a **zero** of f if $(\varphi f)(x) = 0$. Such zeros will be studied in Chapter IX.

Go back to A as a subring of B. Elements $x_1, \ldots, x_n \in B$ are called **algebraically independent** over A if the evaluation map

$$f \mapsto f(x)$$

is injective. Equivalently, we could say that if $f \in A[X]$ is a polynomial and $f(x) = 0$, then $f = 0$; in other words, there are no non-trivial polynomial relations among $x_1, \ldots, x_n$ over A.

Example. It is not known if e and π are algebraically independent over the rationals. It is not even known if $e + \pi$ is rational.

We now come to the notion of degree for several variables. By the **degree** of a primitive monomial

$$M_{(\nu)}(X) = X_1^{\nu_1} \cdots X_n^{\nu_n}$$

we shall mean the integer $|\nu| = \nu_1 + \cdots + \nu_n$ (which is $\geqq 0$).

A polynomial

$$aX_1^{\nu_1} \cdots X_n^{\nu_n} \qquad (a \in A)$$

will be called a **monomial** (not necessarily primitive).

If $f(X)$ is a polynomial in $A[X]$ written as

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

then either $f = 0$, in which case we say that its degree is $-\infty$, or $f \neq 0$, and then we define the **degree** of f to be the maximum of the degrees of the monomials $M_{(\nu)}(X)$ such that $a_{(\nu)} \neq 0$. (Such monomials are said to **occur** in the polynomial.) We note that the degree of f is 0 if and only if

$$f(X) = a_0 X_1^0 \cdots X_n^0$$

for some $a_0 \in A$, $a_0 \neq 0$. We also write this polynomial simply $f(X) = a_0$, i.e. writing 1 instead of

$$X_1^0 \cdots X_n^0,$$

in other words, we identify the polynomial with the constant a_0.

Note that a polynomial $f(X_1, \ldots, X_n)$ in n variables can be viewed as a polynomial in X_n with coefficients in $A[X_1, \ldots, X_{n-1}]$ (if $n \geqq 2$). Indeed, we can write

$$f(X) = \sum_{j=0}^{d_n} f_j(X_1, \ldots, X_{n-1})X_n^j,$$

where f_j is an element of $A[X_1, \ldots, X_{n-1}]$. By the **degree of f in X_n** we shall mean its degree when viewed as a polynomial in X_n with coefficients in $A[X_1, \ldots, X_{n-1}]$. One sees easily that if this degree is d, then d is the largest integer occurring as an exponent of X_n in a monomial

$$a_{(\nu)}X_1^{\nu_1} \cdots X_n^{\nu_n}$$

with $a_{(\nu)} \neq 0$. Similarly, we define the degree of f in each variable X_i ($i = 1, \ldots, n$).

The degree of f in each variable is of course usually different from its degree (which is sometimes called the **total degree** if there is need to prevent ambiguity). For instance,

$$X_1^3X_2 + X_2^2$$

has total degree 4, and has degree 3 in X_1 and 2 in X_2.

As a matter of notation, we shall often abbreviate "degree" by "deg."

For each integer $d \geqq 0$, given a polynomial f, let $f^{(d)}$ be the sum of all monomials occurring in f and having degree d. Then

$$f = \sum_d f^{(d)}.$$

Suppose $f \neq 0$. We say that f is **homogeneous** of degree d if $f = f^{(d)}$; thus f can be written in the form

$$f(X) = \sum a_{(\nu)}X_1^{\nu_1} \cdots X_n^{\nu_n} \qquad \text{with} \qquad \nu_1 + \cdots + \nu_n = d \qquad \text{if} \quad a_{(\nu)} \neq 0.$$

We shall leave it as an exercise to prove that *a non-zero polynomial f in n variables over A is homogeneous of degree d if and only if, for every set of $n + 1$ algebraically independent elements $u, t_1, \ldots, t_n$ over A we have*

$$f(ut_1, \ldots, ut_n) = u^d f(t_1, \ldots, t_n).$$

We note that if f, g are homogeneous of degree d, e respectively, and $fg \neq 0$, then fg is homogeneous of degree $d + e$. If $d = e$ and $f + g \neq 0$, then $f + g$ is homogeneous of degree d.

Remark. In view of the isomorphism

$$A[X_1, \ldots, X_n] \approx A[t_1, \ldots, t_n]$$

between the polynomial ring in n variables and a ring generated over A by n

algebraically independent elements, we can apply all the terminology we have defined for polynomials, to elements of $A[t_1, \dots, t_n]$. Thus we can speak of the degree of an element in $A[t]$, and the rules for the degree of a product or sum hold. In fact, we shall also call elements of $A[t]$ polynomials in (t). Algebraically independent elements will also be called **variables** (or independent variables), and any distinction which we make between $A[X]$ and $A[t]$ is more psychological than mathematical.

Suppose next that A is entire. By what we know of polynomials in one variable and induction, it follows that $A[X_1, \dots, X_n]$ is entire. In particular, suppose f has degree d and g has degree e. Write

$$f = f^{(d)} + \text{terms of lower degree},$$

$$g = g^{(e)} + \text{terms of lower degree}.$$

Then $fg = f^{(d)}g^{(e)} + \text{terms of lower degree}$, and if $fg \neq 0$ then $f^{(d)}g^{(e)} \neq 0$. Thus we find:

$$\deg(fg) = \deg f + \deg g,$$

$$\deg(f + g) \leqq \max(\deg f, \deg g).$$

We are now finished with the basic terminology of polynomials. We end this section by indicating how the construction of polynomials is actually a special case of another construction which is used in other contexts. Interested readers can skip immediately to Chapter IV, giving further important properties of polynomials. See also Exercise 33 of Chapter XIII for harmonic polynomials.

The group ring or monoid ring

Let A be a commutative ring. Let G be a monoid, written multiplicatively.

Let $A[G]$ be the set of all maps $\alpha: G \to A$ such that $\alpha(x) = 0$ for almost all $x \in G$. We define addition in $A[G]$ to be the ordinary addition of mappings into an abelian (additive) group. If α, $\beta \in A[G]$, we define their product $\alpha\beta$ by the rule

$$(\alpha\beta)(z) = \sum_{xy=z} \alpha(x)\beta(y).$$

The sum is taken over all pairs (x, y) with x, $y \in G$ such that $xy = z$. This sum is actually finite, because there is only a finite number of pairs of elements $(x, y) \in G \times G$ such that $\alpha(x)\beta(y) \neq 0$. We also see that $(\alpha\beta)(t) = 0$ for almost all t, and thus belongs to our set $A[G]$.

The axioms for a ring are trivially verified. We shall carry out the proof of associativity as an example. Let α, β, $\gamma \in A[G]$. Then

$$\begin{aligned}((\alpha\beta)\gamma)(z) &= \sum_{xy=z} (\alpha\beta)(x)\gamma(y)\\ &= \sum_{xy=z}\left[\sum_{uv=x} \alpha(u)\beta(v)\right]\gamma(y)\\ &= \sum_{xy=z}\left[\sum_{uv=x} \alpha(u)\beta(v)\gamma(y)\right]\\ &= \sum_{\substack{(u,v,y)\\ uvy=z}} \alpha(u)\beta(v)\gamma(y),\end{aligned}$$

this last sum being taken over all triples $(u\ v,\ y)$ whose product is z. This last sum is now symmetric, and if we had computed $(\alpha(\beta\gamma))(z)$, we would have found this sum also. This proves associativity.

The unit element of $A[G]$ is the function δ such that $\delta(e) = 1$ and $\delta(x) = 0$ for all $x \in G$, $x \neq e$. It is trivial to verify that $\alpha = \delta\alpha = \alpha\delta$ for all $\alpha \in A[G]$.

We shall now adopt a notation which will make the structure of $A[G]$ clearer. Let $a \in A$ and $x \in G$. We denote by $a \cdot x$ (and sometimes also by ax) the function whose value at x is a, and whose value at y is 0 if $y \neq x$. Then an element $\alpha \in A[G]$ can be written as a sum

$$\alpha = \sum_{x\in G} \alpha(x)\cdot x.$$

Indeed, if $\{a_x\}_{x\in G}$ is a set of elements of A almost all of which are 0, and we set

$$\beta = \sum_{x\in G} a_x \cdot x,$$

then for any $y \in G$ we have $\beta(y) = a_y$ (directly from the definitions). This also shows that a given element α admits a unique expression as a sum $\sum a_x \cdot x$.

With our present notation, multiplication can be written

$$\left(\sum_{x\in G} a_x\cdot x\right)\left(\sum_{y\in G} b_y\cdot y\right) = \sum_{x,y} a_x b_y \cdot xy$$

and addition can be written

$$\sum_{x\in G} a_x\cdot x + \sum_{x\in G} b_x\cdot x = \sum_{x\in G} (a_x + b_x)\cdot x,$$

which looks the way we want it to look. Note that the unit element of $A[G]$ is simply $1 \cdot e$.

We shall now see that we can embed both A and G in a natural way in $A[G]$.

Let $\varphi_0\colon G \to A[G]$ be the map given by $\varphi_0(x) = 1 \cdot x$. It is immediately verified that φ_0 is a multiplicative monoid-homomorphism, and is in fact injective, i.e. an embedding.

Let $f_0\colon A \to A[G]$ be the map given by

$$f_0(a) = a \cdot e.$$

It is immediately verified that f_0 is a ring-homomorphism, and is also an embedding. Thus we view A as a subring of $A[G]$. One calls $A[G]$ the **monoid ring** or **monoid algebra** of G over A, or the **group algebra** if G is a group.

Examples. When G is a finite group and $A = k$ is a field, then the group ring $k[G]$ will be studied in Chapter XVIII.

Polynomial rings are special cases of the above construction. In n variables, consider a multiplicative free abelian group of rank n. Let $X_1, \ldots, X_n$ be generators. Let G be the multiplicative subset consisting of elements $X_1^{\nu_1} \cdots X_n^{\nu_n}$ with $\nu_i \geqq 0$ for all i. Then G is a monoid, and the reader can verify at once that $A[G]$ is just $A[X_1, \ldots, X_n]$.

As a matter of notation we usually omit the dot in writing an element of the ring $A[G]$, so we write simply $\sum a_x x$ for such an element.

More generally, let $I = \{i\}$ be an infinite family of indices, and let S be the free abelian group with free generators X_i, written multiplicatively. Then we can form the polynomial ring $A[X]$ by taking the monoid to consist of products

$$M_{(\nu)}(X) = \prod_{i \in I} X_i^{\nu_i},$$

where of course all but a finite number of exponents ν_i are equal to 0. If A is a subring of the commutative ring B, and S is a subset of B, then we shall also use the following notation. Let $\nu \colon S \to \mathbf{N}$ be a mapping which is 0 except for a finite number of elements of S. We write

$$M_{(\nu)}(S) = \prod_{x \in S} x^{\nu(x)}.$$

Thus we get polynomials in infinitely many variables. One interesting example of the use of such polynomials will occur in Artin's proof of the existence of the algebraic closure of a field, cf. Chapter V, Theorem 2.5.

We now consider the evaluation and reduction homomorphisms in the present context of monoids.

Proposition 3.1. *Let $\varphi \colon G \to G'$ be a homomorphism of monoids. Then there exists a unique homomorphism $h \colon A[G] \to A[G']$ such that $h(x) = \varphi(x)$ for all $x \in G$ and $h(a) = a$ for all $a \in A$.*

Proof. In fact, let $\alpha = \sum a_x x \in A[G]$. Define

$$h(\alpha) = \sum a_x \varphi(x).$$

Then h is immediately verified to be a homomorphism of abelian groups, and $h(x) = \varphi(x)$. Let $\beta = \sum b_y y$. Then

$$h(\alpha\beta) = \sum_z \left(\sum_{xy=z} a_x b_y \right) \varphi(z).$$

We get $h(\alpha\beta) = h(\alpha)h(\beta)$ immediately from the hypothesis that $\varphi(xy) =$

$\varphi(x)\varphi(y)$. If e is the unit element of G, then by definition $\varphi(e) = e'$, so Proposition 3.1 follows.

Proposition 3.2. *Let G be a monoid and let $f: A \to B$ be a homomorphism of commutative rings. Then there is a unique homomorphism*

$$h: A[G] \to B[G]$$

such that

$$h\left(\sum_{x \in G} a_x x\right) = \sum_{x \in G} f(a_x)x.$$

Proof. Since every element of $A[G]$ has a unique expression as a sum $\sum a_x x$, the formula giving h gives a well-defined map from $A[G]$ into $B[G]$. This map is obviously a homomorphism of abelian groups. As for multiplication, let

$$\alpha = \sum a_x x \quad \text{and} \quad \beta = \sum b_y y.$$

Then

$$\begin{aligned} h(\alpha\beta) &= \sum_{z \in G} f\left(\sum_{xy=z} a_x b_y\right) z \\ &= \sum_{z \in G} \sum_{xy=z} f(a_x)f(b_y)z \\ &= h(\alpha)h(\beta). \end{aligned}$$

We have trivially $h(1) = 1$, so h is a ring-homomorphism, as was to be shown.

Observe that viewing A as a subring of $A[G]$, the restriction of h to A is the homomorphism f itself. In other words, if e is the unit element of G, then

$$h(ae) = f(a)e.$$

§4. LOCALIZATION

We continue to let A be a commutative ring.

By a **multiplicative subset** of A we shall mean a submonoid of A (viewed as a multiplicative monoid according to **RI 2**). In other words, it is a subset S containing 1, and such that, if $x, y \in S$, then $xy \in S$.

We shall now construct the **quotient ring of** A **by** S, also known as the **ring of fractions of** A **by** S.

We consider pairs (a, s) with $a \in A$ and $s \in S$. We define a relation

$$(a, s) \sim (a', s')$$

between such pairs, by the condition that there exists an element $s_1 \in S$ such that

$$s_1(s'a - sa') = 0.$$

It is then trivially verified that this is an equivalence relation, and the equivalence class containing a pair (a, s) is denoted by a/s. The set of equivalence classes is denoted by $S^{-1}A$.

Note that if $0 \in S$, then $S^{-1}A$ has precisely one element, namely 0/1.

We define a multiplication in $S^{-1}A$ by the rule

$$(a/s)(a'/s') = aa'/ss'.$$

It is trivially verified that this is well defined. This multiplication has a unit element, namely 1/1, and is clearly associative.

We define an addition in $S^{-1}A$ by the rule

$$\frac{a}{s} + \frac{a'}{s'} = \frac{s'a + sa'}{ss'}.$$

It is trivially verified that this is well defined. As an example, we give the proof in detail. Let $a_1/s_1 = a/s$, and let $a_1'/s_1' = a'/s'$. We must show that

$$(s_1'a_1 + s_1a_1')/s_1s_1' = (s'a + sa')/ss'.$$

There exist $s_2, s_3 \in S$ such that

$$s_2(sa_1 - s_1a) = 0,$$

$$s_3(s'a_1' - s_1'a') = 0.$$

We multiply the first equation by $s_3s's_1'$ and the second by s_2ss_1. We then add, and obtain

$$s_2s_3[s's_1'(sa_1 - s_1a) + ss_1(s'a_1' - s_1'a')] = 0.$$

By definition, this amounts to what we want to show, namely that there exists an element of S (e.g. s_2s_3) which when multiplied with

$$ss'(s_1'a_1 + s_1a_1') - s_1s_1'(s'a + sa')$$

yields 0.

We observe that given $a \in A$ and $s, s' \in S$ we have

$$a/s = s'a/s's.$$

Thus this aspect of the elementary properties of fractions still remains true in our present general context.

Finally, it is also trivially verified that our two laws of composition on $S^{-1}A$ define a ring structure.

We let

$$\varphi_S \colon A \to S^{-1}A$$

be the map such that $\varphi_S(a) = a/1$. Then one sees at once that φ_S is a ring-homomorphism. Furthermore, every element of $\varphi_S(S)$ is invertible in $S^{-1}A$ (the inverse of $s/1$ is $1/s$).

Let $\mathcal{C}$ be the category whose objects are ring-homomorphisms

$$f: A \to B$$

such that for every $s \in S$, the element $f(s)$ is invertible in B. If $f: A \to B$ and $f': A \to B'$ are two objects of $\mathcal{C}$, a morphism g of f into f' is a homomorphism

$$g: B \to B'$$

making the diagram commutative:

$$\begin{tikzcd} A \arrow[rr, "f"] \arrow[dr, "f'"'] & & B \arrow[dl, "g"] \\ & B' & \end{tikzcd}$$

We contend that φ_S is a universal object in this category $\mathcal{C}$.

Proof. Suppose that $a/s = a'/s'$, or in other words that the pairs (a, s) and (a', s') are equivalent. There exists $s_1 \in S$ such that

$$s_1(s'a - sa') = 0.$$

Let $f: A \to B$ be an object of $\mathcal{C}$. Then

$$f(s_1)[f(s')f(a) - f(s)f(a')] = 0.$$

Multiplying by $f(s_1)^{-1}$, and then by $f(s')^{-1}$ and $f(s)^{-1}$, we obtain

$$f(a)f(s)^{-1} = f(a')f(s')^{-1}.$$

Consequently, we can define a map

$$h: S^{-1}A \to B$$

such that $h(a/s) = f(a)f(s)^{-1}$, for all $a/s \in S^{-1}A$. It is trivially verified that h is a homomorphism, and makes the usual diagram commutative. It is also trivially verified that such a map h is unique, and hence that φ_S is the required universal object.

Let A be an entire ring, and let S be a multiplicative subset which does not contain 0. *Then*

$$\varphi_S: A \to S^{-1}A$$

is injective.

Indeed, by definition, if $a/1 = 0$ then there exists $s \in S$ such that $sa = 0$, and hence $a = 0$.

The most important cases of a multiplicative set S are the following:

1. Let A be a commutative ring, and let S be the set of invertible elements of A (i.e. the set of units). Then S is obviously multiplicative, and is denoted frequently by A^*. If A is a field, then A^* is the multiplicative group of non-zero elements of A. In that case, $S^{-1}A$ is simply A itself.

2. Let A be an entire ring, and let S be the set of non-zero elements of A. Then S is a multiplicative set, and $S^{-1}A$ is then a field, called the **quotient field** or the **field of fractions**, of A. It is then customary to identify A as a subset of $S^{-1}A$, and we can write

$$a/s = s^{-1}a$$

for $a \in A$ and $s \in S$.

We have seen in §3 that when A is an entire ring, then $A[X_1, \dots, X_n]$ is also entire. If K is the quotient field of A, the quotient field of $A[X_1, \dots, X_n]$ is denoted by $K(X_1, \dots, X_n)$. An element of $K(X_1, \dots, X_n)$ is called a **rational function**. A rational function can be written as a quotient $f(X)/g(X)$ where f, g are polynomials. If $(b_1, \dots, b_n)$ is in $K^{(n)}$, and a rational function admits an expression as a quotient f/g such that $g(b) \neq 0$, then we say that the rational function is **defined** at (b). From general localization properties, we see that when this is the case, we can substitute (b) in the rational function to get a value $f(b)/g(b)$.

3. A ring A is called a **local ring** if it is commutative and has a unique maximal ideal. If A is a local ring and $\mathfrak{m}$ is its maximal ideal, and $x \in A$, $x \notin \mathfrak{m}$, then x is a unit (otherwise x generates a proper ideal, not contained in $\mathfrak{m}$, which is impossible). Let A be a ring and $\mathfrak{p}$ a prime ideal. Let S be the complement of $\mathfrak{p}$ in A. Then S is a multiplicative subset of A, and $S^{-1}A$ is denoted by $A_{\mathfrak{p}}$. It is a local ring (cf. Exercise 3) and is called **the local ring of A at** $\mathfrak{p}$. Cf. the examples of principal rings, and Exercises 15, 16.

Let S be a multiplicative subset of A. Denote by $J(A)$ the set of ideals of A. Then we can define a map

$$\psi_S \colon J(A) \to J(S^{-1}A);$$

namely we let $\psi_S(\mathfrak{a}) = S^{-1}\mathfrak{a}$ be the subset of $S^{-1}A$ consisting of all fractions a/s with $a \in \mathfrak{a}$ and $s \in S$. The reader will easily verify that $S^{-1}\mathfrak{a}$ is an $S^{-1}A$-ideal, and that ψ_S is a homomorphism for both the additive and multiplicative monoid structures on the set of ideals $J(A)$. Furthermore, ψ_S also preserves intersections and inclusions; in other words, for ideals $\mathfrak{a}$, $\mathfrak{b}$ of A we have:

$$S^{-1}(\mathfrak{a} + \mathfrak{b}) = S^{-1}\mathfrak{a} + S^{-1}\mathfrak{b}, \qquad S^{-1}(\mathfrak{a}\mathfrak{b}) = (S^{-1}\mathfrak{a})(S^{-1}\mathfrak{b}),$$

$$S^{-1}(\mathfrak{a} \cap \mathfrak{b}) = S^{-1}\mathfrak{a} \cap S^{-1}\mathfrak{b}.$$

As an example, we prove this last relation. Let $x \in \mathfrak{a} \cap \mathfrak{b}$. Then x/s is in $S^{-1}\mathfrak{a}$ and also in $S^{-1}\mathfrak{b}$, so the inclusion is trivial. Conversely, suppose we have an element of $S^{-1}A$ which can be written as $a/s = b/s'$ with $a \in \mathfrak{a}$, $b \in \mathfrak{b}$, and $s, s' \in S$. Then there exists $s_1 \in S$ such that

$$s_1 s' a = s_1 s b,$$

and this element lies in both $\mathfrak{a}$ and $\mathfrak{b}$. Hence

$$a/s = s_1 s'a/s_1 s's$$

lies in $S^{-1}(\mathfrak{a} \cap \mathfrak{b})$, as was to be shown.

§5. PRINCIPAL AND FACTORIAL RINGS

Let A be an entire ring. An element $a \neq 0$ is called **irreducible** if it is not a unit, and if whenever one can write $a = bc$ with $b \in A$ and $c \in A$ then b or c is a unit.

Let $a \neq 0$ be an element of A and assume that the principal ideal (a) is prime. Then a is irreducible. Indeed, if we write $a = bc$, then b or c lies in (a), say b. Then we can write $b = ad$ with some $d \in A$, and hence $a = acd$. Since A is entire, it follows that $cd = 1$, in other words, that c is a unit.

The converse of the preceding assertion is not always true. We shall discuss under which conditions it is true. An element $a \in A$, $a \neq 0$, is said to have a **unique factorization into irreducible elements** if there exists a unit u and there exist irreducible elements p_i $(i = 1, \dots, r)$ in A such that

$$a = u \prod_{i=1}^{r} p_i,$$

and if given two factorizations into irreducible elements,

$$a = u \prod_{i=1}^{r} p_i = u' \prod_{j=1}^{s} q_j,$$

we have $r = s$, and after a permutation of the indices i, we have $p_i = u_i q_i$ for some unit u_i in A, $i = 1, \dots, r$.

We note that if p is irreducible and u is a unit, then up is also irreducible, so we must allow multiplication by units in a factorization. In the ring of integers $\mathbf{Z}$, the ordering allows us to select a representative irreducible element (a prime number) out of two possible ones differing by a unit, namely $\pm p$, by selecting the positive one. This is, of course, impossible in more general rings.

Taking $r = 0$ above, we adopt the convention that a unit of A has a factorization into irreducible elements.

A ring is called **factorial** (or a **unique factorization ring**) if it is entire and if every element $\neq 0$ has a unique factorization into irreducible elements. We shall prove below that a principal entire ring is factorial.

Let A be an entire ring and $a, b \in A$, $ab \neq 0$. We say that a **divides** b and write $a|b$ if there exists $c \in A$ such that $ac = b$. We say that $d \in A$, $d \neq 0$, is a **greatest common divisor (g.c.d.)** of a and b if $d|a$, $d|b$, and if any element e of A, $e \neq 0$, which divides both a and b also divides d.

Proposition 5.1. *Let A be a principal entire ring and a, $b \in A$, a, $b \neq 0$. Let $(a) + (b) = (c)$. Then c is a greatest common divisor of a and b.*

Proof. Since b lies in the ideal (c), we can write $b = xc$ for some $x \in A$, so that $c|b$. Similarly, $c|a$. Let d divide both a and b, and write $a = dy$, $b = dz$ with $y, z \in A$. Since c lies in (a, b) we can write

$$c = wa + tb$$

with some $w, t \in A$. Then $c = w\,dy + t\,dz = d(wy + tz)$, whence $d|c$, and our proposition is proved.

Theorem 5.2. *Let A be a principal entire ring. Then A is factorial.*

Proof. We first prove that every non-zero element of A has a factorization into irreducible elements. Let S be the set of principal ideals $\neq 0$ whose generators do not have a factorization into irreducible elements, and suppose S is not empty. Let (a_1) be in S. Consider an ascending chain

$$(a_1) \subsetneqq (a_2) \subsetneqq \cdots \subsetneqq (a_n) \subsetneqq \cdots$$

of ideals in S. We contend that such a chain cannot be infinite. Indeed, the union of such a chain is an ideal of A, which is principal, say equal to (a). The generator a must already lie in some element of the chain, say (a_n), and then we see that $(a_n) \subset (a) \subset (a_n)$, whence the chain stops at (a_n). Hence S is inductively ordered, and has a maximal element (a). Therefore any ideal of A containing (a) and $\neq (a)$ has a generator admitting a factorization.

We note that a cannot be irreducible (otherwise it has a factorization), and hence we can write $a = bc$ with neither b nor c equal to a unit. But then $(b) \neq (a)$ and $(c) \neq (a)$ and hence both b, c admit factorizations into irreducible elements. The product of these factorizations is a factorization for a, contradicting the assumption that S is not empty.

To prove uniqueness, we first remark that if p is an irreducible element of A and $a, b \in A$, $p|ab$, then $p|a$ or $p|b$. *Proof*: If $p \nmid a$, then the g.c.d. of p, a is 1 and hence we can write

$$1 = xp + ya$$

with some $x, y \in A$. Then $b = bxp + yab$, and since $p|ab$ we conclude that $p|b$.

Suppose that a has two factorizations

$$a = p_1 \cdots p_r = q_1 \cdots q_s$$

into irreducible elements. Since p_1 divides the product farthest to the right, p_1 divides one of the factors, which we may assume to be q_1 after renumbering these factors. Then there exists a unit u_1 such that $q_1 = u_1 p_1$. We can now cancel p_1 from both factorizations and get

$$p_2 \cdots p_r = u_1 q_2 \cdots q_s.$$

The argument is completed by induction.

We could call two elements a, $b \in A$ equivalent if there exists a unit u such that $a = bu$. Let us select one irreducible element p out of each equivalence class belonging to such an irreducible element, and let us denote by P the set of such representatives. Let $a \in A$, $a \neq 0$. Then there exists a unit u and integers $v(p) \geqq 0$, equal to 0 for almost all $p \in P$ such that

$$a = u \prod_{p \in P} p^{v(p)}.$$

Furthermore, the unit u and the integers $v(p)$ are uniquely determined by a. We call $v(p)$ the **order** of a at p, also written $\text{ord}_p\, a$.

If A is a factorial ring, then an irreducible element p generates a prime ideal (p). Thus in a factorial ring, an irreducible element will also be called a **prime element**, or simply a **prime**.

We observe that one can define the notion of **least common multiple** (l.c.m.) of a finite number of non-zero elements of A in the usual manner: If

$$a_1, \ldots, a_n \in A$$

are such elements, we define a l.c.m. for these elements to be any $c \in A$ such that for all primes p of A we have

$$\text{ord}_p\, c = \max_i \text{ord}_p\, a_i.$$

This element c is well defined up to a unit.

If $a, b \in A$ are non-zero elements, we say that a, b are **relatively prime** if the g.c.d. of a and b is a unit.

Example. The ring of integers $\mathbf{Z}$ is factorial. Its group of units consists of 1 and -1. It is natural to take as representative prime element the positive prime element (what is called a prime number) p from the two possible choices p and $-p$. Similarly, we shall show later that the ring of polynomials in one variable over a field is factorial, and one selects representatives for the prime elements to be the irreducible polynomials with leading coefficient 1.

Examples. It will be proved in Chapter IV that if R is a factorial ring, then the polynomial ring $R[X_1, \ldots, X_n]$ in n variables is factorial. In particular, if k is a field, then the polynomial ring $k[X_1, \ldots, X_n]$ is factorial. Note that $k[X_1]$ is a principal ring, but for $n \geqq 2$, the ring $k[X_1, \ldots, X_n]$ is not principal.

In Exercise 5 you will prove that the localization of a factorial ring is factorial.

In Chapter IV, §9 we shall prove that the power series ring $k[[X_1, \ldots, X_n]]$ is factorial. This result is a special case of the more general statement that a regular local ring is factorial, but we do not define regular local rings in this book. You can look them up in books on commutative algebra. I recommend:

H. MATSUMURA, *Commutative Algebra*, second edition, Benjamin-Cummings, New York, 1980

H. MATSUMURA, *Commutative Rings*, Cambridge University Press, Cambridge, UK, 1986

Examples from algebraic and complex geometry. Roughly speaking, regular local rings arise in the following context of algebraic or complex geometry. Consider the ring of regular functions in the neighborhood of some point on a complex or algebraic manifold. This ring is regular. A typical example is the ring of convergent power series in a neighborhood of 0 in $\mathbf{C}^n$. In Chapter IV, we shall prove some results on power series which give some algebraic background for those analytic theories, and which are used in proving the factoriality of rings of power series, convergent or not.

Conversely to the above examples, singularities in geometric theories may give rise to examples of non-factoriality. We give examples using notions which are sufficiently basic so that readers should have encountered them in more elementary courses.

Examples of non-factorial rings. Let k be a field, and let x be a variable over k. Let $R = k[x^2, x^3]$. Then R is not factorial (proof?). The ring R may be viewed as the ring of regular functions on the curve $y^2 = x^3$, which has a singularity at the origin, as you can see by drawing its real graph.

Let R be the set of all numbers of the form $a + b\sqrt{-5}$, where $a, b \in \mathbf{Z}$. Then the only units of R are ± 1, and the elements $3, 2 + \sqrt{-5}, 2 - \sqrt{-5}$ are irreducible elements, giving rise to a non-unique factorization

$$3^2 = (2 + \sqrt{-5})(2 - \sqrt{-5}).$$

(Do Exercise 10.) Here the non-factoriality is not due to singularities but due to a non-trivial ideal class group of R, which is a Dedekind ring. For a definition see the exercises of Chapter III, or go straight to my book *Algebraic Number Theory*, for instance.

As Trotter once pointed out (*Math. Monthly*, April 1988), the relation

$$\sin^2 x = (1 + \cos x)(1 - \cos x)$$

may be viewed as a non-unique factorization in the ring of trigonometric polynomials $\mathbf{R}[\sin x, \cos x]$, generated over $\mathbf{R}$ by the functions $\sin x$ and $\cos x$. This ring is a subring of the ring of all functions, or of all differentiable functions. See Exercise 11.

EXERCISES

We let A denote a commutative ring.

1. Suppose that $1 \neq 0$ in A. Let S be a multiplicative subset of A not containing 0. Let $\mathfrak{p}$ be a maximal element in the set of ideals of A whose intersection with S is empty. Show that $\mathfrak{p}$ is prime.

2. Let $f: A \to A'$ be a surjective homomorphism of rings, and assume that A is local, $A' \neq 0$. Show that A' is local.

3. Let $\mathfrak{p}$ be a prime ideal of A. Show that $A_{\mathfrak{p}}$ has a unique maximal ideal, consisting of all elements a/s with $a \in \mathfrak{p}$ and $s \notin \mathfrak{p}$.

4. Let A be a principal ring and S a multiplicative subset with $0 \notin S$. Show that $S^{-1}A$ is principal.

5. Let A be a factorial ring and S a multiplicative subset with $0 \notin S$. Show that $S^{-1}A$ is factorial, and that the prime elements of $S^{-1}A$ are of the form up with primes p of A such that $(p) \cap S$ is empty, and units u in $S^{-1}A$.

6. Let A be a factorial ring and p a prime element. Show that the local ring $A_{(p)}$ is principal.

7. Let A be a principal ring and $a_1, \ldots, a_n$ non-zero elements of A. Let $(a_1, \ldots, a_n) = (d)$. Show that d is a greatest common divisor for the a_i $(i = 1, \ldots, n)$.

8. Let p be a prime number, and let A be the ring $\mathbf{Z}/p^r\mathbf{Z}$ (r = integer $\geqq 1$). Let G be the group of units in A, i.e. the group of integers prime to p, modulo p^r. Show that G is cyclic, except in the case when

$$p = 2, \qquad r \geqq 3,$$

in which case it is of type $(2, 2^{r-2})$. [*Hint*: In the general case, show that G is the product of a cyclic group generated by $1 + p$, and a cyclic group of order $p - 1$. In the exceptional case, show that G is the product of the group $\{\pm 1\}$ with the cyclic group generated by the residue class of 5 mod 2^r.]

9. Let i be the complex number $\sqrt{-1}$. Show that the ring $\mathbf{Z}[i]$ is principal, and hence factorial. What are the units?

10. Let D be an integer $\geqq 1$, and let R be the set of all elements $a + b\sqrt{-D}$ with $a, b \in \mathbf{Z}$.
    (a) Show that R is a ring.
    (b) Using the fact that complex conjugation is an automorphism of $\mathbf{C}$, show that complex conjugation induces an automorphism of R.
    (c) Show that if $D \geqq 2$ then the only units in R are ± 1.
    (d) Show that 3, $2 + \sqrt{-5}$, $2 - \sqrt{-5}$ are irreducible elements in $\mathbf{Z}[\sqrt{-5}]$.

11. Let R be the ring of trigonometric polynomials as defined in the text. Show that R consists of all functions f on $\mathbf{R}$ which have an expression of the form

$$f(x) = a_0 + \sum_{m=1}^{n} (a_m \cos mx + b_m \sin mx),$$

where a_0, a_m, b_m are real numbers. Define the **trigonometric degree** $\deg_{\text{tr}}(f)$ to be the maximum of the integers r, s such that a_r, $b_s \neq 0$. Prove that

$$\deg_{\text{tr}}(fg) = \deg_{\text{tr}}(f) + \deg_{\text{tr}}(g).$$

Deduce from this that R has no divisors of 0, and also deduce that the functions $\sin x$ and $1 - \cos x$ are irreducible elements in that ring.

12. Let P be the set of positive integers and R the set of functions defined on P with values in a commutative ring K. Define the sum in R to be the ordinary addition of functions, and define the **convolution product** by the formula

$$(f * g)(m) = \sum_{xy=m} f(x)g(y),$$

where the sum is taken over all pairs (x, y) of positive integers such that $xy = m$.
(a) Show that R is a commutative ring, whose unit element is the function δ such that $\delta(1) = 1$ and $\delta(x) = 0$ if $x \neq 1$.
(b) A function f is said to be **multiplicative** if $f(mn) = f(m)f(n)$ whenever m, n are relatively prime. If f, g are multiplicative, show that $f * g$ is multiplicative.
(c) Let μ be the **Möbius function** such that $\mu(1) = 1$, $\mu(p_1 \cdots p_r) = (-1)^r$ if $p_1, \ldots, p_r$ are distinct primes, and $\mu(m) = 0$ if m is divisible by p^2 for some prime p. Show that $\mu * \varphi_1 = \delta$, where φ_1 denotes the constant function having value 1. [*Hint*: Show first that μ is multiplicative, and then prove the assertion for prime powers.] The Möbius inversion formula of elementary number theory is then nothing else but the relation $\mu * \varphi_1 * f = f$.

Dedekind rings

Prove the following statements about a Dedekind ring $\mathfrak{o}$. To simplify terminology, by an **ideal** we shall mean non-zero ideal unless otherwise specified. We let K denote the quotient field of $\mathfrak{o}$.

13. Every ideal is finitely generated. [*Hint*: Given an ideal $\mathfrak{a}$, let $\mathfrak{b}$ be the fractional ideal such that $\mathfrak{a}\mathfrak{b} = \mathfrak{o}$. Write $1 = \sum a_i b_i$ with $a_i \in \mathfrak{a}$ and $b_i \in \mathfrak{b}$. Show that $\mathfrak{a} = (a_1, \ldots, a_n)$.]

14. Every ideal has a factorization as a product of prime ideals, uniquely determined up to permutation.

15. Suppose $\mathfrak{o}$ has only one prime ideal $\mathfrak{p}$. Let $t \in \mathfrak{p}$ and $t \notin \mathfrak{p}^2$. Then $\mathfrak{p} = (t)$ is principal.

16. Let $\mathfrak{o}$ be any Dedekind ring. Let $\mathfrak{p}$ be a prime ideal. Let $\mathfrak{o}_\mathfrak{p}$ be the local ring at $\mathfrak{p}$. Then $\mathfrak{o}_\mathfrak{p}$ is Dedekind and has only one prime ideal.

17. As for the integers, we say that $\mathfrak{a}|\mathfrak{b}$ ($\mathfrak{a}$ **divides** $\mathfrak{b}$) if there exists an ideal $\mathfrak{c}$ such that $\mathfrak{b} = \mathfrak{a}\mathfrak{c}$. Prove:
(a) $\mathfrak{a}|\mathfrak{b}$ if and only if $\mathfrak{b} \subset \mathfrak{a}$.
(b) Let $\mathfrak{a}$, $\mathfrak{b}$ be ideals. Then $\mathfrak{a} + \mathfrak{b}$ is their greatest common divisor. In particular, $\mathfrak{a}$, $\mathfrak{b}$ are relatively prime if and only if $\mathfrak{a} + \mathfrak{b} = \mathfrak{o}$.

18. Every prime ideal $\mathfrak{p}$ is maximal. (Remember, $\mathfrak{p} \neq 0$ by convention.) In particular, if $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$ are distinct primes, then the Chinese remainder theorem applies to their powers $\mathfrak{p}_1^{r_1}, \ldots, \mathfrak{p}_n^{r_n}$. Use this to prove:

19. Let $\mathfrak{a}$, $\mathfrak{b}$ be ideals. Show that there exists an element $c \in K$ (the quotient field of $\mathfrak{o}$) such that $c\mathfrak{a}$ is an ideal relatively prime to $\mathfrak{b}$. In particular, every ideal class in $\operatorname{Pic}(\mathfrak{o})$ contains representative ideals prime to a given ideal.

For a continuation, see Exercise 7 of Chapter VII; Chapter III, Exercise 11–13.

CHAPTER III

Modules

Although this chapter is logically self-contained and prepares for future topics, in practice readers will have had some acquaintance with vector spaces over a field. We generalize this notion here to modules over rings. It is a standard fact (to be reproved) that a vector space has a basis, but for modules this is not always the case. Sometimes they do; most often they do not. We shall look into cases where they do.

For examples of modules and their relations to those which have a basis, the reader should look at the comments made at the end of §4.

§1. BASIC DEFINITIONS

Let A be a ring. A **left module** over A, or a left A-module M is an abelian group, usually written additively, together with an operation of A on M (viewing A as a multiplicative monoid by **RI 2**), such that, for all $a, b \in A$ and $x, y \in M$ we have

$$(a + b)x = ax + bx \quad \text{and} \quad a(x + y) = ax + ay.$$

We leave it as an exercise to prove that $a(-x) = -(ax)$ and that $0x = 0$.ˋ By definition of an operation, we have $1x = x$.

In a similar way, one defines a **right A-module**. We shall deal only with left A-modules, unless otherwise specified, and hence call these simply **A-modules**, or even **modules** if the reference is clear.

S. Lang, *Algebra*

Let M be an A-module. By a **submodule** N of M we mean an additive subgroup such that $AN \subset N$. Then N is a module (with the operation induced by that of A on M).

Examples

We note that A is a module over itself.

Any commutative group is a **Z**-module.

An additive group consisting of 0 alone is a module over any ring.

Any left ideal of A is a module over A.

Let J be a two-sided ideal of A. Then the factor ring A/J is actually a module over A. If $a \in A$ and $x + J$ is a coset of J in A, then one defines the operation to be $a(x + J) = ax + J$. The reader can verify at once that this defines a module structure on A/J. More general, if M is a module and N a submodule, we shall define the factor module below. Thus if L is a left ideal of A, then A/L is also a module. For more examples in this vein, see §4.

A module over a field is called a **vector space**. Even starting with vector spaces, one is led to consider modules over rings. Indeed, let V be a vector space over the field K. The reader no doubt already knows about linear maps (which will be recalled below systematically). Let R be the ring of all linear maps of V into itself. Then V is a module over R. Similarly, if $V = K^n$ denotes the vector space of (vertical) n-tuples of elements of K, and R is the ring of $n \times n$ matrices with components in K, then V is a module over R. For more comments along these lines, see the examples at the end of §2.

Let S be a non-empty set and M an A-module. Then the set of maps $\mathrm{Map}(S, M)$ is an A-module. We have already noted previously that it is a commutative group, and for $f \in \mathrm{Map}(S, M)$, $a \in A$ we define af to be the map such that $(af)(s) = af(s)$. The axioms for a module are then trivially verified. For further examples, see the end of this section.

For the rest of this section, we deal with a fixed ring A, and hence may omit the prefix A-.

Let A be an *entire* ring and let M be an A-module. We define the **torsion submodule** M_{tor} to be the subset of elements $x \in M$ such that there exists $a \in A$, $a \neq 0$ such that $ax = 0$. It is immediately verified that M_{tor} is a submodule. Its structure in an important case will be determined in §7.

Let $\mathfrak{a}$ be a left ideal, and M a module. We define $\mathfrak{a}M$ to be the set of all elements

$$a_1x_1 + \cdots + a_nx_n$$

with $a_i \in \mathfrak{a}$ and $x_i \in M$. It is obviously a submodule of M. If $\mathfrak{a}$, $\mathfrak{b}$ are left ideals, then we have associativity, namely

$$\mathfrak{a}(\mathfrak{b}M) = (\mathfrak{a}\mathfrak{b})M.$$

We also have some obvious distributivities, like $(\mathfrak{a} + \mathfrak{b})M = \mathfrak{a}M + \mathfrak{b}M$. If N, N' are submodules of M, then $\mathfrak{a}(N + N') = \mathfrak{a}N + \mathfrak{a}N'$.

Let M be an A-module, and N a submodule. We shall define a module structure on the factor group M/N (for the additive group structure). Let $x + N$ be a coset of N in M, and let $a \in A$. We define $a(x + N)$ to be the coset $ax + N$. It is trivial to verify that this is well defined (i.e. if y is in the same coset as x, then ay is in the same coset as ax), and that this is an operation of A on M/N satisfying the required condition, making M/N into a module, called the **factor module** of M by N.

By a **module-homomorphism** one means a map

$$f: M \to M'$$

of one module into another (over the same ring A), which is an additive group-homomorphism, and such that

$$f(ax) = af(x)$$

for all $a \in A$ and $x \in M$. It is then clear that the collection of A-modules is a category, whose morphisms are the module-homomorphisms usually also called homomorphisms for simplicity, if no confusion is possible. If we wish to refer to the ring A, we also say that f is an **A-homomorphism**, or also that it is an **A-linear map**.

If M is a module, then the identity map is a homomorphism. For any module M', the map $\zeta : M \to M'$ such that $\zeta(x) = 0$ for all $x \in M$ is a homomorphism, called **zero**.

In the next section, we shall discuss the homomorphisms of a module into itself, and as a result we shall give further examples of modules which arise in practice. Here we continue to tabulate the translation of basic properties of groups to modules.

Let M be a module and N a submodule. We have the canonical additive group-homomorphism

$$f: M \to M/N$$

and one verifies trivially that it is a module-homomorphism.

Equally trivially, one verifies that f is universal in the category of homomorphisms of M whose kernel contains N.

If $f: M \to M'$ is a module-homomorphism, then its kernel and image are submodules of M and M' respectively (trivial verification).

Let $f: M \to M'$ be a homomorphism. By the **cokernel** of f we mean the factor module $M'/\mathrm{Im}\, f = M'/f(M)$. One may also mean the canonical homomorphism

$M' \to M'/f(M)$ rather than the module itself. The context should make clear which is meant. Thus the cokernel is a factor module of M'.

Canonical homomorphisms discussed in Chapter I, §3 apply to modules *mutatis mutandis*. For the convenience of the reader, we summarise these homomorphisms:

Let N, N' be two submodules of a module M. Then $N + N'$ is also a submodule, and we have an isomorphism

$$N/(N \cap N') \approx (N + N')/N'.$$

If $M \supset M' \supset M''$ are modules, then

$$(M/M'')/(M'/M'') \approx M/M'.$$

If $f : M \to M'$ is a module-homomorphism, and N' is a submodule of M', then $f^{-1}(N')$ is a submodule of M and we have a canonical injective homomorphism

$$\bar{f} : M/f^{-1}(N') \to M'/N'.$$

If f is surjective, then $\bar{f}$ is a module-isomorphism.

The proofs are obtained by verifying that all homomorphisms which appeared when dealing with abelian groups are now A-homomorphisms of modules. We leave the verification to the reader.

As with groups, we observe that a module-homomorphism which is bijective is a module-isomorphism. Here again, the proof is the same as for groups, adding only the observation that the inverse map, which we know is a group-isomorphism, actually is a module-isomorphism. Again, we leave the verification to the reader.

As with abelian groups, we define a sequence of module-homomorphisms

$$M' \xrightarrow{f} M \xrightarrow{g} M''$$

to be **exact** if $\operatorname{Im} f = \operatorname{Ker} g$. We have an exact sequence associated with a submodule N of a module M, namely

$$0 \to N \to M \to M/N \to 0,$$

the map of N into M being the inclusion, and the subsequent map being the canonical map. The notion of exactness is due to Eilenberg-Steenrod.

If a homomorphism $u : N \to M$ is such that

$$0 \to N \xrightarrow{u} M$$

is exact, then we also say that u is a **monomorphism** or an **embedding**. Dually, if

$$N \xrightarrow{u} M \to 0$$

is exact, we say that u is an **epimorphism**.

Algebras

There are some things in mathematics which satisfy all the axioms of a ring except for the existence of a unit element. We gave the example of $L^1(\mathbf{R})$ in Chapter II, §1. There are also some things which do not satisfy associativity, but satisfy distributivity. For instance let R be a ring, and for $x, y \in R$ define the **bracket product**

$$[x, y] = xy - yx.$$

Then this bracket product is not associative in most cases when R is not commutative, but it satisfies the distributive law.

Examples. A typical example is the ring of differential operators with C^∞ coefficients, operating on the ring of C^∞ functions on an open set in $\mathbf{R}^n$. The bracket product

$$[D_1, D_2] = D_1 \circ D_2 - D_2 \circ D_1$$

of two differential operators is again a differential operator. In the theory of Lie groups, the tangent space at the origin also has such a bracket product.

Such considerations lead us to define a more general notion than a ring. Let A be a commutative ring. Let E, F be modules. By a **bilinear map**

$$g: E \times E \to F$$

we mean a map such that given $x \in E$, the map $y \mapsto g(x, y)$ is A-linear, and given $y \in E$, the map $x \mapsto g(x, y)$ is A-linear. By an A-**algebra** we mean a module together with a bilinear map $g: E \times E \to E$. We view such a map as a law of composition on E. But in this book, unless otherwise specified, we shall assume that our algebras are associative and have a unit element.

Aside from the examples already mentioned, we note that the group ring $A[G]$ (or monoid ring when G is a monoid) is an A-algebra, also called the **group** (or **monoid**) **algebra**. Actually the group algebra can be viewed as a special case of the following situation.

Let $f: A \to B$ be a ring-homomorphism such that $f(A)$ is contained in the center of B, i.e., $f(a)$ commutes with every element of B for every $a \in A$. Then we may view B as an A-module, defining the operation of A on B by the map

$$(a, b) \mapsto f(a)b$$

for all $a \in A$ and $b \in B$. The axioms for a module are trivially satisfied, and the multiplicative law of composition $B \times B \to B$ is clearly bilinear (i.e., A-bilinear). In this book, unless otherwise specified, by an **algebra** over A, we shall always mean a ring-homomorphism as above. We say that the algebra is **finitely generated** if B is finitely generated as a ring over $f(A)$.

Several examples of modules over a polynomial algebra or a group algebra will be given in the next section, where we also establish the language of representations.

§2. THE GROUP OF HOMOMORPHISMS

Let A be a ring, and let X, X' be A-modules. We denote by $\mathrm{Hom}_A(X', X)$ the set of A-homomorphisms of X' into X. Then $\mathrm{Hom}_A(X', X)$ is an abelian group, the law of addition being that of addition for mappings into an abelian group.

If A *is commutative* then we can make $\mathrm{Hom}_A(X', X)$ into an A-module, by defining af for $a \in A$ and $f \in \mathrm{Hom}_A(X', X)$ to be the map such that

$$(af)(x) = af(x).$$

The verification that the axioms for an A-module are satisfied is trivial. However, if A is not commutative, then we view $\mathrm{Hom}_A(X', X)$ simply as an abelian group.

We also view Hom_A as a functor. It is actually a functor of two variables, contravariant in the first and covariant in the second. Indeed, let Y be an A-module, and let

$$X' \xrightarrow{f} X$$

be an A-homomorphism. Then we get an induced homomorphism

$$\mathrm{Hom}_A(f, Y) : \mathrm{Hom}_A(X, Y) \to \mathrm{Hom}_A(X', Y)$$

(reversing the arrow!) given by

$$g \mapsto g \circ f.$$

This is illustrated by the following sequence of maps:

$$X' \xrightarrow{f} X \xrightarrow{g} Y.$$

The fact that $\mathrm{Hom}_A(f, Y)$ is a homomorphism is simply a rephrasing of the property $(g_1 + g_2) \circ f = g_1 \circ f + g_2 \circ f$, which is trivially verified. If $f = \mathrm{id}$, then composition with f acts as an identity mapping on g, i.e. $g \circ \mathrm{id} = g$.

If we have a sequence of A-homomorphisms

$$X' \to X \to X'',$$

then we get an induced sequence

$$\mathrm{Hom}_A(X', Y) \leftarrow \mathrm{Hom}_A(X, Y) \leftarrow \mathrm{Hom}_A(X'', Y).$$

Proposition 2.1. *A sequence*

$$X' \xrightarrow{\lambda} X \to X'' \to 0$$

is exact if and only if the sequence

$$\mathrm{Hom}_A(X', Y) \leftarrow \mathrm{Hom}_A(X, Y) \leftarrow \mathrm{Hom}_A(X'', Y) \leftarrow 0$$

is exact for all Y.

Proof. This is an important fact, whose proof is easy. For instance, suppose the first sequence is exact. If $g : X'' \to Y$ is an A-homomorphism, its image in $\mathrm{Hom}_A(X, Y)$ is obtained by composing g with the surjective map of X on X''. If this composition is 0, it follows that $g = 0$ because $X \to X''$ is surjective. As another example, consider a homomorphism $g : X \to Y$ such that the composition

$$X' \xrightarrow{\lambda} X \xrightarrow{g} Y$$

is 0. Then g vanishes on the image of λ. Hence we can factor g through the factor module,

$$\begin{tikzcd}
& X/\mathrm{Im}\,\lambda \arrow[dr] & \\
X \arrow[ur] \arrow[rr, "g"'] & & Y
\end{tikzcd}$$

Since $X \to X''$ is surjective, we have an isomorphism

$$X/\mathrm{Im}\,\lambda \leftrightarrow X''.$$

Hence we can factor g through X'', thereby showing that the kernel of

$$\mathrm{Hom}_A(X', Y) \leftarrow \mathrm{Hom}_A(X, Y)$$

is contained in the image of

$$\mathrm{Hom}_A(X, Y) \leftarrow \mathrm{Hom}_A(X'', Y).$$

The other conditions needed to verify exactness are left to the reader. So is the converse.

We have a similar situation with respect to the second variable, but then the functor is covariant. Thus if X is fixed, and we have a sequence of A-homomorphisms

$$Y' \to Y \to Y'',$$

then we get an induced sequence

$$\mathrm{Hom}_A(X, Y') \to \mathrm{Hom}_A(X, Y) \to \mathrm{Hom}_A(X, Y'').$$

Proposition 2.2. *A sequence*

$$0 \to Y' \to Y \to Y'',$$

is exact if and only if

$$0 \to \mathrm{Hom}_A(X, Y') \to \mathrm{Hom}_A(X, Y) \to \mathrm{Hom}_A(X, Y'')$$

is exact for all X.

The verification will be left to the reader. It follows at once from the definitions.

We note that to say that

$$0 \to Y' \to Y$$

is exact means that Y' is embedded in Y, i.e. is isomorphic to a submodule of Y. A homomorphism into Y' can be viewed as a homomorphism into Y if we have $Y' \subset Y$. This corresponds to the injection

$$0 \to \mathrm{Hom}_A(X, Y') \to \mathrm{Hom}_A(X, Y).$$

Let Mod(A) and Mod(B) be the categories of modules over rings A and B, and let $F\colon \mathrm{Mod}(A) \to \mathrm{Mod}(B)$ be a functor. One says that F is **exact** if F transforms exact sequences into exact sequences. We see that the Hom functor in either variable need not be exact if the other variable is kept fixed. In a later section, we define conditions under which exactness is preserved.

Endomorphisms. Let M be an A-module. From the relations

$$(g_1 + g_2) \circ f = g_1 \circ f + g_2 \circ f$$

and its analogue on the right, namely

$$g \circ (f_1 + f_2) = g \circ f_1 + g \circ f_2,$$

and the fact that there is an identity for composition, namely id_M, we conclude that $\mathrm{Hom}_A(M, M)$ is a ring, the multiplication being defined as composition of mappings. If n is an integer $\geqq 1$, we can write f^n to mean the iteration of f with itself n times, and define f^0 to be id. According to the general definition of endomorphisms in a category, we also write $\mathrm{End}_A(M)$ instead of $\mathrm{Hom}_A(M, M)$, and we call $\mathrm{End}_A(M)$ the ring of **endomorphisms**.

Since an A-module M is an abelian group, we see that $\mathrm{Hom}_{\mathbf{Z}}(M, M)$ (= set of group-homomorphisms of M into itself) is a ring, and that we could have defined an operation of A on M to be a ring-homomorphism $A \to \mathrm{Hom}_{\mathbf{Z}}(M, M)$.

Let A be *commutative*. Then M is a module over $\mathrm{End}_A(M)$. If R is a subring of $\mathrm{End}_A(M)$ then M is *a fortiori* a module over R. More generally, let R be a ring and let $\rho\colon R \to \mathrm{End}_A(M)$ be a ring homomorphism. Then ρ is called a **representation** of R on M. This occurs especially if $A = K$ is a field. The linear algebra of representations of a ring will be discussed in Part III, in several contexts, mostly finite-dimensional. Infinite-dimensional examples occur in analysis, but then the representation theory mixes algebra with analysis, and thus goes beyond the level of this course.

Example. Let K be a field and let V be a vector space over K. Let $D\colon V \to V$ be an endomorphism (K-linear map). For every polynomial $P(X) \in K[X]$, $P(X) = \sum a_i X^i$ with $a_i \in K$, we can define

$$P(D) = \sum a_i D^i \colon V \to V$$

as an endomorphism of V. The association $P(X) \mapsto P(D)$ gives a representation

$$\rho \colon K[X] \to \mathrm{End}_K(V),$$

which makes V into a $K[X]$-module. It will be shown in Chapter IV that $K[X]$ is a principal ring. In §7 we shall give a general structure theorem for modules over principal rings, which will be applied to the above example in the context of linear algebra for finite-dimensional vector spaces in Chapter XIV, §3. Readers acquainted with basic linear algebra from an undergraduate course may wish to read Chapter XIV already at this point.

Examples for infinite-dimensional vector spaces occur in analysis. For instance, let V be the vector space of complex-valued C^∞ functions on $\mathbf{R}$. Let $D = d/dt$ be the derivative (if t is the variable). Then $D \colon V \to V$ is a linear map, and $\mathbf{C}[X]$ has the representation $\rho \colon \mathbf{C}[X] \to \mathrm{End}_{\mathbf{C}}(V)$ given by $P \mapsto P(D)$. A similar situation exists in several variables, when we let V be the vector space of C^∞ functions in n variables on an open set of $\mathbf{R}^n$. Then we let $D_i = \partial/\partial t_i$ be the partial derivative with respect to the i-th variable ($i = 1, \dots, n$). We obtain a representation

$$\rho \colon \mathbf{C}[X_1, \dots, X_n] \to \mathrm{End}_{\mathbf{C}}(V)$$

such that $\rho(X_i) = D_i$.

Example. Let H be a Hilbert space and let A be a bounded hermitian operator on A. Then one considers the homomorphism $\mathbf{R}[X] \to \mathbf{R}[A] \subset \mathrm{End}(H)$, from the polynomial ring into the algebra of endomorphisms of H, and one extends this homomorphism to the algebra of continuous functions on the spectrum of A. *Cf.* my *Real and Functional Analysis*, Springer Verlag, 1993.

Representations form a category as follows. We define a **morphism** of a representation $\rho \colon R \to \mathrm{End}_A(M)$ into a representation $\rho' \colon R \to \mathrm{End}_A(M')$, or in other words a **homomorphism of one representation of R to another**, to be an A-module homomorphism $h \colon M \to M'$ such that the following diagram is commutative for every $\alpha \in R$:

$$\begin{tikzcd}
M \arrow[r, "h"] \arrow[d, "\rho(\alpha)"'] & M' \arrow[d, "\rho'(\alpha)"] \\
M \arrow[r, "h"'] & M'
\end{tikzcd}$$

In the case when h is an isomorphism, then we may replace the above diagram by the commutative diagram

$$\begin{tikzcd}
 & \mathrm{End}_A(M) \arrow[d, "{[h]}"] \\
R \arrow[ur, "\rho"] \arrow[dr, "\rho'"'] & \\
 & \mathrm{End}_A(M')
\end{tikzcd}$$

where the symbol $[h]$ denotes conjugation by h, i.e. for $f \in \operatorname{End}_A(M)$ we have $[h]f = h \circ f \circ h^{-1}$.

Representations: from a monoid to the monoid algebra. Let G be a monoid. By a **representation of** G on an A-module M, we mean a homomorphism $\rho\colon G \to \operatorname{End}_A(M)$ of G into the multiplicative monoid of $\operatorname{End}_A(M)$. Then we may extend ρ to a homomorphism of the monoid algebra

$$A[G] \to \operatorname{End}_A(M),$$

by letting

$$\rho\Bigl(\sum_{x \in G} a_x x\Bigr) = \sum_{x \in G} a_x \rho(x).$$

It is immediately verified that this extension of ρ to $A[G]$ is a ring homomorphism, coinciding with the given ρ on elements of G.

Examples: modules over a group ring. The next examples will follow a certain pattern associated with groups of automorphisms. Quite generally, suppose we have some category of objects, and to each object K there is associated an abelian group $F(K)$, functorially with respect to isomorphisms. This means that if $\sigma\colon K \to K'$ is an isomorphism, then there is an associated isomorphism $F(\sigma)\colon F(K) \to F(K')$ such that $F(\mathrm{id}_K) = \mathrm{id}_{K'}$ and $F(\sigma\tau) = F(\sigma) \circ F(\tau)$. Then the group of automorphisms $\operatorname{Aut}(K)$ of an object operates on $F(K)$; that is, we have a natural homomorphism

$$\operatorname{Aut}(K) \to \operatorname{Aut}(F(K)) \quad \text{given by} \quad \sigma \mapsto F(\sigma).$$

Let $G = \operatorname{Aut}(K)$. Then $F(K)$ (written additively) can be made into a module over the group ring $\mathbf{Z}[G]$ as above. Given an element $\alpha = \sum a_\sigma \sigma \in \mathbf{Z}[G]$, with $a_\sigma \in \mathbf{Z}$, and an element $x \in F(K)$, we define

$$\alpha x = \sum a_\sigma F(\sigma) x.$$

The conditions defining a module are trivially satisfied. We list several concrete cases from mathematics at large, so there are no holds barred on the terminology.

Let K be a number field (i.e. a finite extension of the rational numbers). Let G be its group of automorphisms. Associated with K we have the following objects:

the ring of algebraic integers $\mathfrak{o}_K$;

the group of units $\mathfrak{o}_K^*$;

the group of ideal classes $C(K)$;

the group of roots of unity $\boldsymbol{\mu}(K)$.

Then G operates on each of those objects, and one problem is to determine the structure of these objects as $\mathbf{Z}[G]$-modules. Already for cyclotomic fields this determination gives rise to substantial theories and to a number of unsolved problems.

Suppose that K is a Galois extension of k with Galois group G (see Chapter VI). Then we may view K itself as a module over the group ring $k[G]$. In Chapter VI, §13 we shall prove that K is isomorphic to $k[G]$ as module over $k[G]$ itself.

In topology, one considers a space X_0 and a finite covering X. Then $\mathrm{Aut}(X/X_0)$ operates on the homology of X, so this homology is a module over the group ring.

With more structure, suppose that X is a projective non-singular variety, say over the complex numbers. Then to X we can associate:

the group of divisor classes (Picard group) $\mathrm{Pic}(X)$;

in a given dimension, the group of cycle classes or Chow group $\mathrm{CH}^p(X)$;

the ordinary homology of X;

the sheaf cohomology in general.

If X is defined over a field K finitely generated over the rationals, we can associate a fancier cohomology defined algebraically by Grothendieck, and functorial with respect to the operation of Galois groups.

Then again all these objects can be viewed as modules over the group ring of automorphism groups, and major problems of mathematics consist in determining their structure. I direct the reader here to two surveys, which contain extensive bibliographies.

[CCFT 91] P. Cassou-Nogues, T. Chinburg, A. Fröhlich, M. J. Taylor, L-functions and Galois modules, in *L-functions and Arithmetic* J. Coates and M. J. Taylor (eds.), *Proceedings of the Durham Symposium* July 1989, *London Math, Soc. Lecture Note Series* **153**, Cambridge University Press (1991), pp. 75-139

[La 82] S. Lang, Units and class groups in number theory and algebraic geometry, *Bull. AMS* **Vol. 6 No. 3** (1982), pp. 253-316

§3. DIRECT PRODUCTS AND SUMS OF MODULES

Let A be a ring. Let $\{M_i\}_{i\in I}$ be a family of modules. We defined their direct product as abelian groups in Chapter I, §9. Given an element $(x_i)_{i\in I}$ of the direct product, and $a \in A$, we define $a(x_i) = (ax_i)$. In other words, we multiply by an element a componentwise. Then the direct product $\prod M_i$ is an A-module. The reader will verify at once that it is also a **direct product** in the category of A-modules.

Similarly, let

$$M = \bigoplus_{i \in I} M_i$$

be their direct sum as abelian groups. We define on M a structure of A-module: If $(x_i)_{i \in I}$ is an element of M, i.e. a family of elements $x_i \in M_i$ such that $x_i = 0$ for almost all i, and if $a \in A$, then we define

$$a(x_i)_{i \in I} = (ax_i)_{i \in I},$$

that is we define multiplication by a componentwise. It is trivially verified that this is an operation of A on M which makes M into an A-module. If one refers back to the proof given for the existence of direct sums in the category of abelian groups, one sees immediately that this proof now extends in the same way to show that M is a direct sum of the family $\{M_i\}_{i \in I}$ as A-modules. (For instance, the map

$$\lambda_j : M_j \to M$$

such that $\lambda_j(x)$ has j-th component equal to x and i-th component equal to 0 for $i \neq j$ is now seen to be an A-homomorphism.)

This direct sum is a **coproduct in the category of A-modules**. Indeed, the reader can verify at once that given a family of A-homomorphisms $\{f_i : M_i \to N\}$, the map f defined as in the proof for abelian groups is also an A-isomorphism and has the required properties. See Proposition 7.1 of Chapter I.

When I is a finite set, there is a useful criterion for a module to be a direct product.

Proposition 3.1. *Let M be an A-module and n an integer $\geqq 1$. For each $i = 1, \ldots, n$ let $\varphi_i : M \to M$ be an A-homomorphism such that*

$$\sum_{i=1}^{n} \varphi_i = \mathrm{id} \quad \textit{and} \quad \varphi_i \circ \varphi_j = 0 \qquad \textit{if } i \neq j.$$

Then $\varphi_i^2 = \varphi_i$ for all i. Let $M_i = \varphi_i(M)$, and let $\varphi : M \to \prod M_i$ be such that

$$\varphi(x) = (\varphi_1(x), \ldots, \varphi_n(x)).$$

Then φ is an A-isomorphism of M onto the direct product $\prod M_i$.

Proof. For each j, we have

$$\varphi_j = \varphi_j \circ \mathrm{id} = \varphi_j \circ \sum_{i=1}^{n} \varphi_i = \varphi_j \circ \varphi_j = \varphi_j^2,$$

thereby proving the first assertion. It is clear that φ is an A-homomorphism. Let x be in its kernel. Since

$$x = \mathrm{id}(x) = \sum_{i=1}^{n} \varphi_i(x)$$

we conclude that $x = 0$, so φ is injective. Given elements $y_i \in M_i$ for each $i = 1, \dots, n$, let $x = y_1 + \cdots + y_n$. We obviously have $\varphi_j(y_i) = 0$ if $i \neq j$. Hence

$$\varphi_j(x) = y_j$$

for each $j = 1, \dots, n$. This proves that φ is surjective, and concludes the proof of our proposition.

We observe that when I is a finite set, the direct sum and the direct product are equal.

Just as with abelian groups, we use the symbol $\oplus$ to denote direct sum.

Let M be a module over a ring A and let S be a subset of M. By a **linear combination** of elements of S (with coefficients in A) one means a sum

$$\sum_{x \in S} a_x x$$

where $\{a_x\}$ is a set of elements of A, almost all of which are equal to 0. These elements a_x are called the **coefficients** of the linear combination. Let N be the set of all linear combinations of elements of S. Then N is a submodule of M, for if

$$\sum_{x \in S} a_x x \quad \text{and} \quad \sum_{x \in S} b_x x$$

are two linear combinations, then their sum is equal to

$$\sum_{x \in S} (a_x + b_x)x,$$

and if $c \in A$, then

$$c\left(\sum_{x \in S} a_x x\right) = \sum_{x \in S} c a_x x,$$

and these elements are again linear combinations of elements of S. We shall call N the submodule **generated** by S, and we call S a set of **generators** for N. We sometimes write $N = A\langle S \rangle$. If S consists of one element x, the module generated by x is also written Ax, or simply (x), and sometimes we say that (x) is a **principal module**.

A module M is said to be **finitely generated**, or of **finite type**, or **finite** over A, if it has a finite number of generators.

A *subset* S of a module M is said to be **linearly independent** (over A) if whenever we have a linear combination

$$\sum_{x \in S} a_x x$$

which is equal to 0, then $a_x = 0$ for all $x \in S$. If S is linearly independent and if two linear combinations

$$\sum a_x x \quad \text{and} \quad \sum b_x x$$

are equal, then $a_x = b_x$ for all $x \in S$. Indeed, subtracting one from the other yields $\sum (a_x - b_x)x = 0$, whence $a_x - b_x = 0$ for all x. If S is linearly independent we shall also say that its elements are linearly independent. Similarly, a *family* $\{x_i\}_{i \in I}$ of elements of M is said to be linearly independent if whenever we have a linear combination

$$\sum_{i \in I} a_i x_i = 0,$$

then $a_i = 0$ for all i. A subset S (resp. a family $\{x_i\}$) is called **linearly dependent** if it is not linearly independent, i.e. if there exists a relation

$$\sum_{x \in S} a_x x = 0 \quad \text{resp.} \quad \sum_{i \in I} a_i x_i = 0$$

with not all a_x (resp. a_i) $= 0$. **Warning**. Let x be a single element of M which is linearly independent. Then the family $\{x_i\}_{i=1,\dots,n}$ such that $x_i = x$ for all i is linearly dependent if $n > 1$, but the set consisting of x itself is linearly independent.

Let M be an A-module, and let $\{M_i\}_{i \in I}$ be a family of submodules. Since we have inclusion-homomorphisms

$$\lambda_i : M_i \to M$$

we have an induced homomorphism

$$\lambda_* : \bigoplus M_i \to M$$

which is such that for any family of elements $(x_i)_{i \in I}$, all but a finite number of which are 0, we have

$$\lambda_*((x_i)) = \sum_{i \in I} x_i.$$

If λ_* is an isomorphism, then we say that the family $\{M_i\}_{i \in I}$ is a **direct sum decomposition** of M. This is obviously equivalent to saying that every element of M has a unique expression as a sum

$$\sum x_i$$

with $x_i \in M_i$, and almost all $x_i = 0$. By abuse of notation, we also write

$$M = \bigoplus M_i$$

in this case.

If the family $\{M_i\}$ is such that every element of M has *some* expression as a sum $\sum x_i$ (not necessarily unique), then we write $M = \sum M_i$. In any case, if $\{M_i\}$ is an arbitrary family of submodules, the image of the homomorphism λ_* above is a submodule of M, which will be denoted by $\sum M_i$.

If M is a module and N, N' are two submodules such that $N + N' = M$ and $N \cap N' = 0$, then we have a module-isomorphism

$$M \approx N \oplus N',$$

just as with abelian groups, and similarly with a finite number of submodules.

We note, of course, that our discussion of abelian groups is a special case of our discussion of modules, simply by viewing abelian groups as modules over $\mathbf{Z}$. However, it seems usually desirable (albeit inefficient) to develop first some statements for abelian groups, and then point out that they are valid (obviously) for modules in general.

Let M, M', N be modules. Then we have an isomorphism of abelian groups

$$\boxed{\operatorname{Hom}_A(M \oplus M', N) \overset{\approx}{\leftrightarrow} \operatorname{Hom}_A(M, N) \times \operatorname{Hom}_A(M', N),}$$

and similarly

$$\boxed{\operatorname{Hom}_A(N, M \times M') \overset{\approx}{\leftrightarrow} \operatorname{Hom}_A(N, M) \times \operatorname{Hom}_A(N, M').}$$

The first one is obtained as follows. If $f: M \oplus M' \to N$ is a homomorphism, then f induces a homomorphism $f_1 : M \to N$ and a homomorphism $f_2 : M' \to N$ by composing f with the injections of M and M' into their direct sum respectively:

$$M \to M \oplus \{0\} \subset M \oplus M' \xrightarrow{f} N,$$

$$M' \to \{0\} \oplus M' \subset M \oplus M' \xrightarrow{f} N.$$

We leave it to the reader to verify that the association

$$f \mapsto (f_1, f_2)$$

gives an isomorphism as in the first box. The isomorphism in the second box is obtained in a similar way. Given homomorphisms

$$f_1 : N \to M$$

and

$$f_2 : N \to M'$$

we have a homomorphism $f: N \to M \times M'$ defined by

$$f(x) = (f_1(x), f_2(x)).$$

It is trivial to verify that the association

$$(f_1, f_2) \mapsto f$$

gives an isomorphism as in the second box.

Of course, the direct sum and direct product of two modules are isomorphic, but we distinguished them in the notation for the sake of functoriality, and to fit the infinite case, see Exercise 22.

Proposition 3.2. *Let* $0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$ *be an exact sequence of modules. The following conditions are equivalent:*

1. *There exists a homomorphism* $\varphi: M'' \to M$ *such that* $g \circ \varphi = \mathrm{id}$.
2. *There exists a homomorphism* $\psi: M \to M'$ *such that* $\psi \circ f = \mathrm{id}$.

If these conditions are satisfied, then we have isomorphisms:

$$M = \mathrm{Im}\, f \oplus \mathrm{Ker}\, \psi, \qquad M = \mathrm{Ker}\, g \oplus \mathrm{Im}\, \varphi,$$

$$M \approx M' \oplus M''.$$

Proof. Let us write the homomorphisms on the right:

$$\begin{tikzcd} M \arrow[r, shift left, "g"] & M'' \arrow[l, shift left, "\varphi"] \arrow[r] & 0. \end{tikzcd}$$

Let $x \in M$. Then

$$x - \varphi(g(x))$$

is in the kernel of g, and hence $M = \mathrm{Ker}\, g + \mathrm{Im}\, \varphi$.

This sum is direct, for if

$$x = y + z$$

with $y \in \mathrm{Ker}\, g$ and $z \in \mathrm{Im}\, \varphi$, $z = \varphi(w)$ with $w \in M''$, and applying g yields $g(x) = w$. Thus w is uniquely determined by x, and therefore z is uniquely determined by x. Hence so is y, thereby proving the sum is direct.

The arguments concerning the other side of the sequence are similar and will be left as exercises, as well as the equivalence between our conditions. When these conditions are satisfied, the exact sequence of Proposition 3.2 is said to **split**. One also says that ψ **splits** f and φ **splits** g.

Abelian categories

Much in the theory of modules over a ring is arrow-theoretic. In fact, one needs only the notion of kernel and cokernel (factor modules). One can axiomatize the special notion of a category in which many of the arguments are valid, especially the arguments used in this chapter. Thus we give this axiomatization now, although for concreteness, at the beginning of the chapter, we continue to use the language of modules. Readers should strike their own balance when they want to slide into the more general framework.

Consider first a category $\mathfrak{A}$ such that $\mathrm{Mor}(E, F)$ is an abelian group for each pair of objects E, F of $\mathfrak{A}$, satisfying the following two conditions:

AB 1. The law of composition of morphisms is bilinear, and there exists a zero object 0, i.e. such that $\mathrm{Mor}(0, E)$ and $\mathrm{Mor}(E, 0)$ have precisely one element for each object E.

AB 2. Finite products and finite coproducts exist in the category.

Then we say that $\mathfrak{A}$ is an **additive category**.

Given a morphism $E \xrightarrow{f} F$ in $\mathfrak{A}$, we define a **kernel** of f to be a morphism $E' \to E$ such that for all objects X in the category, the following sequence is exact:

$$0 \to \mathrm{Mor}(X, E') \to \mathrm{Mor}(X, E) \to \mathrm{Mor}(X, F).$$

We define a **cokernel** for f to be a morphism $F \to F''$ such that for all objects X in the category, the following sequence is exact:

$$0 \to \mathrm{Mor}(F'', X) \to \mathrm{Mor}(F, X) \to \mathrm{Mor}(E, X).$$

It is immediately verified that kernels and cokernels are universal in a suitable category, and hence uniquely determined up to a unique isomorphism if they exist.

AB 3. Kernels and cokernels exist.

AB 4. If $f: E \to F$ is a morphism whose kernel is 0, then f is the kernel of its cokernel. If $f: E \to F$ is a morphism whose cokernel is 0, then f is the cokernel of its kernel. A morphism whose kernel and cokernel are 0 is an isomorphism.

A category $\mathfrak{A}$ satisfying the above four axioms is called an **abelian category**.

In an abelian caegory, the group of morphisms is usually denoted by Hom, so for two objects E, F we write

$$\mathrm{Mor}(E, F) = \mathrm{Hom}(E, F).$$

The morphisms are usually called **homomorphisms**. Given an exact sequence

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

we say that M' is a **subobject** of M, or that the homomorphism of M' into M is a **monomorphism**. Dually, in an exact sequence

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

we say that M'' is a **quotient object** of M, or that the homomorphism of M to M'' is an **epimorphism**, instead of saying that it is surjective as in the category of modules. Although it is convenient to think of modules and abelian groups to construct proofs, usually such proofs will involve only arrow-theoretic arguments, and will therefore apply to any abelian category. However, all the abelian categories we shall meet in this book will have elements, and the kernels and cokernels will be defined in a natural fashion, close to those for modules, so readers may restrict their attention to these concrete cases.

Examples of abelian categories. Of course, modules over a ring form an abelian category, the most common one. Finitely generated modules over a Noetherian ring form an abelian category, to be studied in Chapter X.

Let k be a field. We consider pairs (V, A) consisting of a finite-dimensional vector space V over k, and an endomorphism $A\colon V \to V$. By a **homomorphism (morphism)** of such pairs $f\colon (V, A) \to (W, B)$ we mean a k-homomorphism $f\colon V \to W$ such that the following diagram is commutative:

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "A"'] & W \arrow[d, "B"] \\ V \arrow[r, "f"'] & W \end{tikzcd}$$

It is routinely verified that such pairs and the above defined morphisms form an abelian category. Its elements will be studied in Chapter XIV.

Let k be a field and let G be a group. Let $\mathrm{Mod}_k(G)$ be the category of finite-dimensional vector spaces V over k, with an operation of G on V, i.e. a homomorphism $G \to \mathrm{Aut}_k(V)$. A homomorphism (morphism) in that category is a k-homomorphism $f\colon V \to W$ such that $f(ax) = af(x)$ for all $x \in V$ and $a \in G$. It is immediate that $\mathrm{Mod}_k(G)$ is an abelian category. This category will be studied especially in Chapter XVIII.

In Chapter XX, §1 we shall consider the category of complexes of modules over a ring. This category of complexes is an abelian category.

In topology and differential geometry, the category of vector bundles over a topological space is an abelian category.

Sheaves of abelian groups over a topological space form an abelian category, which will be defined in Chapter XX, §6.

§4. FREE MODULES

Let M be a module over a ring A and let S be a subset of M. We shall say that S is a **basis** of M if S is not empty, if S generates M, and if S is linearly independent. If S is a basis of M, then in particular $M \neq \{0\}$ if $A \neq \{0\}$ and every element of M has a unique expression as a linear combination of elements of S. Similarly, let $\{x_i\}_{i \in I}$ be a non-empty family of elements of M. We say that it is a **basis** of M if it is linearly independent and generates M.

If A is a ring, then as a module over itself, A admits a basis, consisting of the unit element 1.

Let I be a non-empty set, and for each $i \in I$, let $A_i = A$, viewed as an A-module. Let

$$F = \bigoplus_{i \in I} A_i.$$

Then F admits a basis, which consists of the elements e_i of F whose i-th component is the unit element of A_i, and having all other components equal to 0.

By a **free** module we shall mean a module which admits a basis, or the zero module.

Theorem 4.1. *Let A be a ring and M a module over A. Let I be a non-empty set, and let $\{x_i\}_{i \in I}$ be a basis of M. Let N be an A-module, and let $\{y_i\}_{i \in I}$ be a family of elements of N. Then there exists a unique homomorphism $f: M \to N$ such that $f(x_i) = y_i$ for all i.*

Proof. Let x be an element of M. There exists a unique family $\{a_i\}_{i \in I}$ of elements of A such that

$$x = \sum_{i \in I} a_i x_i.$$

We define

$$f(x) = \sum a_i y_i.$$

It is then clear that f is a homomorphism satisfying our requirements, and that it is the unique such, because we must have

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

Corollary 4.2. *Let the notation be as in the theorem, and assume that $\{y_i\}_{i \in I}$ is a basis of N. Then the homomorphism f is an isomorphism, i.e. a module-isomorphism.*

Proof. By symmetry, there exists a unique homomorphism

$$g: N \to M$$